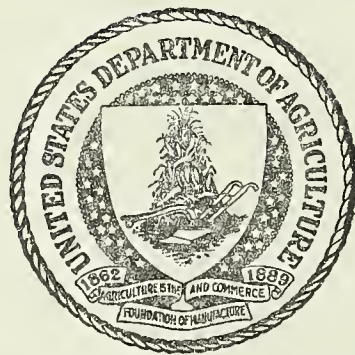


Historic, Archive Document

Do not assume content reflects current scientific knowledge, policies, or practices.

A275.2
F31E

UNITED STATES
DEPARTMENT OF AGRICULTURE
LIBRARY



BOOK NUMBER A275.2
966113 F31E

Discussions

EXTENSION EVALUATION



April 5-23, 1954

Aguas Buenos, Puerto Rico

FEDERAL EXTENSION SERVICE
UNITED STATES DEPARTMENT OF AGRICULTURE
January 1959

DISCUSSION OUTLINE

	<u>Page</u>
Evaluation: Definition, its need, its use	1
Extension objectives	5
Teaching or operating procedure as a basis for evaluation plan	14
Information to be collected	16
From whom to gather information	22
How you will get the information	32
Formulation of questions	33
Tabulation and summarization	40
Interpretation of data	46
Writing your annual report	49
Use of evaluation findings and presentation of data ..	52

E X T E N S I O N E V A L U A T I O N

by

Laurel K. Sabrosky
Extension Analyst

Discussion I. EVALUATION: DEFINITION, ITS NEED, ITS USE.

Definition

This course is a course in Extension Evaluation. It has been apparent for a number of years that in Extension we have been doing a lot of work, going without sleep, running ourselves ragged, trying to accomplish whatever we thought we ought to accomplish; but we have seldom stopped long enough to find out whether or not we have been getting anything worth while done, or whether we were using the best approaches or methods. Whenever somebody did stop long enough to wonder about his success, he usually found he really didn't know how to proceed to find out about it.

The idea and procedure of Extension evaluation developed out of this need. Every worker needs to know of his success or lack of it, and that means that Extension workers do, too. In this course we study and discuss how to find out if we are accomplishing anything. The procedure centers around a rather set outline of steps to be followed. Once you know the basic fundamentals and reasons why, you will have a basic pattern to follow, not one that will fit every circumstance, but one that can be adapted to every circumstance.

This process of finding out your success is called Evaluation.

Evaluation comes from the work value, which means worth, or a quality which makes something useful. It means finding the value of something, expressed in either quantity or quality. It has to do with the determination of the worth of anything. I say "has to do with," because many times evaluation, or a phase of evaluation, does not directly answer the question of worth, but only lays a basis on which you can determine worth.

Extension workers have one over-all objective -- to be of service to their people. That service is usually in the form of teaching. Therefore, when Extension workers wish to carry out evaluation, they wish to find out something which will help them determine the worth of their teaching or other service work.

Evaluation, as we use it, involves a quite definite process of thinking. This process is referred to as scientific thinking; it involves the scientific attitude. Evaluation, of a kind, is always going on, but too often it involves a haphazard thinking process, and not an orderly, scientific one. Scientific thinking proceeds from facts through to a hypothesis or hypotheses and onward to more facts. Haphazard or unscientific thinking invariably omits one or more of the steps -- either jumping to conclusions without facts, (that is, forming hypotheses not based on facts) or gathering stray facts without formulating any hypotheses. Facts in Extension teaching may concern the county situation, the people with whom you work, your own abilities or capabilities, the methods you use, or the number or size of products developed. In answering the question, "Do you know it to be true?", the evaluation process leads you step by step to a valid and reliable answer. Instead of saying "I think the situation is such and such, and therefore we should do this or that," by thinking in an orderly scientific fashion, you can say "I know the situation is such and such, and therefore certain things should be done," or "I do not know what the situation is. I will have to find out before I know what to do next."

Evaluation is often used synonymously with "measurement" and "appraisal." Actually, measurement does not go as far as evaluation. Measurement gets the size or numbers or quantity of something; evaluation goes beyond that and determines whether anything of a certain size, or quantity, or of certain total is of worth or value. (Ex: Room; 4-H enrollment) The Census measures our population; we can take their figures and use them in evaluating what we are trying to do with that population. Appraisal is also finding the value or worth of something. Evaluation and appraisal are interchangeable in general, although I noticed in the dictionary that evaluation was defined as a careful appraisal. When we use the word appraisal in our studies work, we do not connote less careful work; we do usually connote a more general, widespread study of a situation.

Place of Evaluation in The Teaching Plan

Evaluation is a complement to, and a basis for all other phases of Extension work. Evaluation, also, must be in terms of the people with whom you work and the objectives you have in mind, the same as does your plan of work for the year or for any given activity.

In planning work,
the question is:

HOW SHALL I?



In evaluation,
the question is:

DID I?

This may involve:

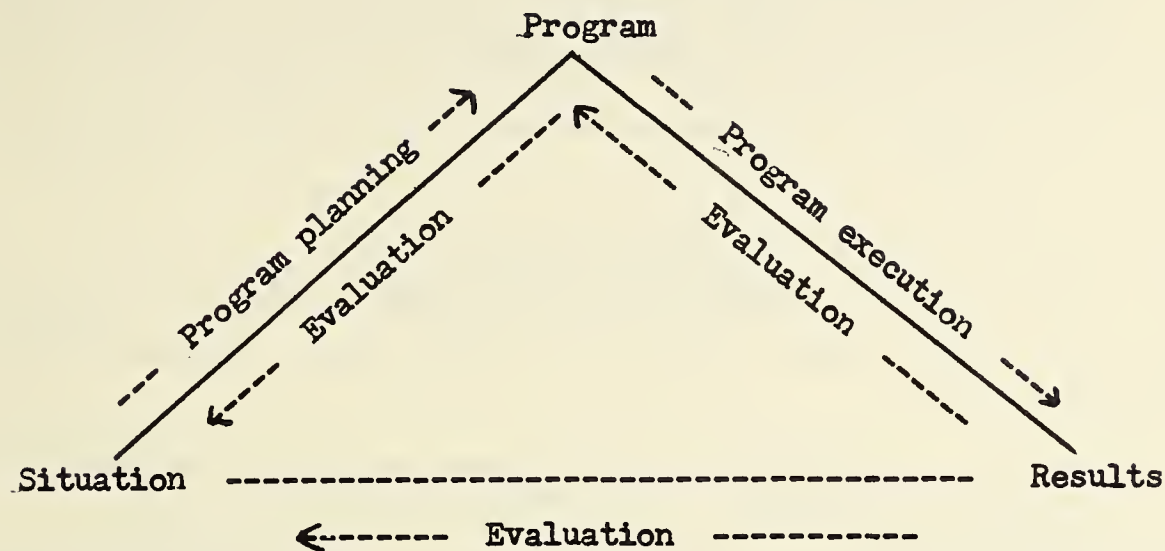
Where were they?
What were they?
Who were they?

←-- From ----

What
made
them
move?

---- To ----->

How
far did
they
go?



Evaluation; Its Purposes

1. It helps us to know if we're doing the right thing! "To validate the hypothesis" ...Tyler.
2. It helps us to determine progress with any given activity or job.
 - (a) It helps us to determine the degree to which we accomplished what we set out to do.
3. It helps us to know more about the people with whom we work - their description, their desires and needs, their interests - and also the people with whom we don't work.
 - (a) It helps us to know the differences between groups, communities, clubs, and individuals.
4. It helps us to judge from impartial and objective evidence the effectiveness of certain methods or activities. This includes program-planning procedures, teaching methods, training, guidance, etc.

All of which leads to:

5. It helps us make necessary modifications in our program.

6. It provides us with information to report to the public - the public including all those involved in providing funds, support in other ways, and just plain interest in the development of their children.
7. It provides us with information - "us" including local leaders, cooperators, and club members as well as professional Extension workers - that gives us satisfaction, a feeling of accomplishment, and confidence in ourselves and in our Service.
8. "... is also a potent method of continued teacher education."
... Tyler.

Purposes of This Course

Stated in rather specific terms, the three primary purposes of the course are to help you -

1. To become acquainted with the process and steps involved in Extension evaluation, and develop some skill in it.
2. To develop a pattern of thinking based on these steps, and therefore become more analytical.
3. To develop an appreciation of the values and purposes of evaluation.

(a) And therefore to see the place of evaluation in the teaching plan.

From the purposes stated above it can be readily seen that it is not the intention of this course to make "studies persons" out of each of the students but rather to help them gain a broad concept of extension program development and teaching, and also to help them visualize methods of systematically appraising their work and the contribution an on-going evaluation plan can make toward increasing the effectiveness of their extension effort.

Evaluation is a continuing, constant procedure. It is not limited to such formal procedures as writing in outline form as you will be doing here this summer. Most of it is informal and a part of the thinking of the worker. However, the most satisfactory way to guide the thinking of the class along the right road is to have you prepare in outline form your evaluation of some phase of your work.

Before we start that procedure, I want to warn you about one of the most common fallacies in the evaluation field. That fallacy is that one must make a study in order to do evaluation; that data must be tabulated; that a report must be written. Again, I repeat that evaluation is a continuing, constant procedure, and by no means limited to such formal procedures as tabulating data and writing reports. Formal studies are but one of the methods of evaluation.

Discussion II. EXTENSION OBJECTIVES

There is one statement and fact that you must keep in your mind all the time and that is: You cannot evaluate until you know what you have been trying to do. You can evaluate only in terms of the objectives of the job. You cannot study the effectiveness of extension work, or of any small part of it, until you are absolutely clear as to the objectives of extension work, or of that small part that you are evaluating. I will stress over and over these 3 weeks that in order to evaluate, you must know your objectives and your people. Your evaluation should tell you if you are reaching your objectives, if you are reaching your people.

This is not more true with evaluation than it is with teaching, of course. It is just as true to say that you cannot teach unless you know the exact objectives of your teaching, and that you cannot teach unless you know and appreciate the population you should reach. WHAT? and WHOM?

Definitions:

An objective can be thought of as a "direction of movement." In what direction do you want your people to move? Which way do you want to go in a county with respect to the dairy enterprise? Toward more dairy cattle, toward better fed cattle, toward greater total milk production, toward higher production per cow, toward greater fat content? Our objectives are often vague, and mean different things to different people. It is as impossible to teach and get anywhere doing so when you are not clear as to the goal of your teaching as it is to go home when you are not clear as to where home is. In each case, the direction and the route have to be known and planned ahead; in both cases, several routes may lead you to the same goal, but yet you must know which route you are going to take and the details of it.

Different educators use different words, such as purpose, aim, goal, and objective. Some of them use them synonymously, others differentiate between them. I intend to use only two of these words when talking about extension teaching -- objectives and goals. In general, objectives mean the direction in which the people will move; goals mean the distance they should go. There may be temporary goals, such as goals for the year, and final goals.

Characteristics of Objectives:

Objectives are desirable pre-determined end-products which one seeks to attain in education. Notice those two adjectives -- (1) desirable, and (2) predetermined. Not all activity has objectives; but all purposeful activity should have objectives -- desirable ones, and, above all, predetermined ones.

A third characteristic of objectives is that they must be clearly defined. So very many, usually too many, of our extension objectives are vague, and mean different things to different people such as citizenship in 4-H. Before you can go ahead and plan the program for the year, the objectives of your work must be stated -- and stated so clearly and accurately that you know exactly where you want your people to go, and that the people with whom you are working know exactly where they hope to go.

When you plan a series of meetings, news stories, circulars, and radio talks on nutrition for the coming year, do you know in exact terms what direction you expect the people to have gone at the end of that program of work? No doubt an objective of your work is to help the people improve their nutrition. That is a vague objective. It involves hundreds, perhaps thousands, of different problems, and projects. The meetings cannot be well planned, the talks cannot be well written or delivered, the letters and circulars cannot be effectively prepared, until you have clearly in your mind exactly in what direction and to what goal you want the people to move. Clarifying objectives keeps you from trying to cover the universe in your work and helps you exclude the less important items or the unnecessary activities.

When we set up a program, we must have clearly in mind what we want to do -- what we expect to accomplish -- what we expect to result from our efforts.

When we set up the objectives for this class, we spent a lot of time, and a lot of discussion, clarifying in our minds exactly what we wanted this class to accomplish. Then, when we finally felt that we knew what we wanted to do, we started to plan exactly what information, and how much of that information, needed to be given to the class members in order that they would move in the direction we want them to move, and as far in that direction as we would want them to move in 3 weeks.

This brings us to an important point in connection with objectives -- and a fourth characteristic, whether of extension teaching or of any other purposeful activity. A GOOD objective in extension work is one that will provide opportunity for a large number of people to move some distance, and in the direction they themselves want to move. Not all people need or desire to move in the same direction or equal distance. In this class you have indicated that you all want to move in the direction you think this class will take you. But you do not all need nor want to move equal distances. Some of you start from one point, some from another. Some of you want to move farther in the distance, some of you wish to move only the minimum amount. We will not go into the particulars here of how to get the people to "want" to move in the right direction -- that is one of the problems of program planning. However, Extension must help people to define the directions in which they desire and need to go -- and then provide assistance to them in traveling in those directions. This is the essence of extension work.

One great problem in making Extension more effective is that of analyzing the stage of readiness. That is deciding on where people are -- and who -- how many, and so forth, at this particular stage -- and then knowing those things -- clarifying the target at which you will aim. It would be most difficult to plan this course if we had no idea of who you people were, and without some idea of where you are along the roads of extension evaluation experience. We feel we have a pretty good idea of who you are, where you are, and of what direction you and we want you to move, and of how far you can and should move in a given length of time.

Necessity for Objectives:

These factors -- these points are all involved in objectives. What do you want to accomplish? You don't go into the kitchen and start mixing ingredients together without knowing what you want to make and knowing how to make it. You don't go into the machine shed and start taking machinery apart without knowing what you are doing and what you want to accomplish, and knowing how to fix the particular trouble. How then can you go ahead and plan a program for nutrition or dairy in your county without knowing what you want to accomplish and what activities will lead you and your people from where they are now to the goal they should reach? What end-product do you want? What ingredients do you need to put together to get a good, successful end-product?

I believe I would not be far wrong if I said that many 4-H Clubs have been organized without any definite idea of the direction the boys and girls involved should take, and how far they should go in that direction each year. I am sure most 4-H Clubs are organized with a definite idea of what direction the CLUB should take and how far the CLUB should go (standard clubs - honor clubs). How can meetings be planned, and how can events be planned and carried out, with no idea, or a vague idea (like better rural boys and girls), as to the direction of movement of the boys and girls, and the goal they should reach? With carefully defined objectives in mind, and a definite goal attainable within a given time, each meeting of a 4-H Club can be planned so that that objective might be reached. Each activity or action within a meeting should help some or all the 4-H Club members turn toward the right direction or move along the right road a little way. And no activity or action should push them back, or over in the other direction a little way.

In evaluation, we try to find out the worth or value of a program, a method, an activity, and so forth. We cannot find out anything about its worth or its value, without knowing the objective it was trying to reach. A meeting is never good or bad, except in terms of the objectives the leader had in mind in calling it. If the people in attendance moved in the direction of the objectives, then there was some goodness in the meeting. If they did not move in that direction at all, it had no goodness. Some of you may find that at some time or other, maybe right here in class, you will be trying to evaluate something for which you realize

you had no concrete, exactly defined objectives. You may have been trying to improve something - but what did you mean by improve? You may have been trying to help the people do something - and whom did you mean by the people?

Again, I will repeat -- one cannot evaluate, or find the worth or success of, anything without knowing the exact objectives of the activity or function being evaluated. Evaluation can be carried out only in terms of the objectives of the activity.

Extension Objectives:

Now, we will discuss more specifically extension objectives. We have all heard expressed over and over, discussed, argued, extension objectives -- what they are, what they ought to be.

The Smith-Lever Act expressed the objective of extension work as, to aid in diffusing among the people of the United States useful and practical information in subjects relating to agriculture and home economics and to encourage the application of the same.

The Smith-Lever Act in no words says that Extension Service should do things for the people -- it says to give them information and encourage them to use it. I do not believe that in any discussion of the objectives of Extension work there is any argument with the basic facts as expressed in the Smith-Lever Act.

Since that Act, there have been many interpretations. Changing conditions have changed the information given, altered opinions as to which subjects relate to agriculture and home economics, brought about new ways of encouraging people, and affected the degrees of encouragement necessary.

Dr. H. C. Ramsower, retired Director of Extension Work in Ohio, stated that "It is not what extension work does for people, but what it does to them that counts. The ultimate objective of extension teaching is to promote the physical, mental, spiritual, and social growth of the individual farmer, his wife, and his children. According to our present point of view, this can best be done by assisting them in analyzing their own problems, in finding solutions for them, and in bringing about active participation in formulating and carrying out the plans necessary to put these solutions into effect."

Dean Bailey, formerly head of the New York State Agricultural College at Cornell, said, in setting forth the objectives of Extension, "(1) To teach those who have a desire for information and (2) to create a desire for information in those who do not yet have the desire."

In these statements, and the many others available down through the years, it is apparent that Extension Service work is considered as a service in education. The word "service" itself may have a misleading connotation to some people. One of the definitions of the word means "performance of labor for the benefit of another." In many cases, extension work is done exactly that way. Another definition of Service is "Act or instance of helping or benefiting another." This seems closer to what was meant when the term "Extension Service" was coined. We are put on our jobs to help or benefit others, and according to the interpretations of our objectives, it means to help or benefit others in educational ways.

Education itself has also had its many definitions and meanings. Dozens of different definitions can be found written by the educators of the past and present. All of them, however, carry the idea if not the word, of developing people. A simple definition is that education is the work of helping people acquire certain skills, understandings, and attitudes. Dr. Kruse, defines education as the product of changes in human behavior. Although neither of these definitely covers acquisition of knowledge as part of educational development, it is usually considered as part of it, not as an end in itself, but as a basic essential for the acquisition of skills, understandings, and attitudes.

Considering the broader Extension objectives, then, together with the broader educational objectives, we find that the objectives of Extension consist of developing people along agricultural and homemaking lines; of helping people acquire knowledge, skills, understandings, and attitudes in the fields of agriculture and homemaking.

The important idea in any of these definitions is that word "people," or man, farmer, homemaker, boy, or girl, or however it is expressed. The work that must be done in extension work is for the purpose of bringing about some change in the man - in the woman - in the boy or girl. In no interpretation of Extension's objectives - in no definition of education - will one find that a better barn, a better soil conservation program, a bigger garden, a better dairy cow, more attractive clothes, or better use of money in the food budget is an objective of Extension Service. The people are the population in whom Extension Service is interested, and changes in the people are the objectives at which Extension Service is aiming. The above-named objectives in terms of equipment, land, or facilities have their place in the extension program of work; they are not ultimate objectives of Extension Service work. Their place will be discussed later.

Classification of Objectives:

Expressing educational objectives in Extension terms, we find that Extension helps people acquire knowledge of:

- Vitamin content of foods
- Feeding practices for animals
- Food budgets
- Amount of fertilizers needed

It helps them acquire skills in:

- Arranging their homes more conveniently
- Pruning a tree
- Putting on fertilizer
- Making a dress

It helps them acquire or change attitudes by explaining the pros and cons of:

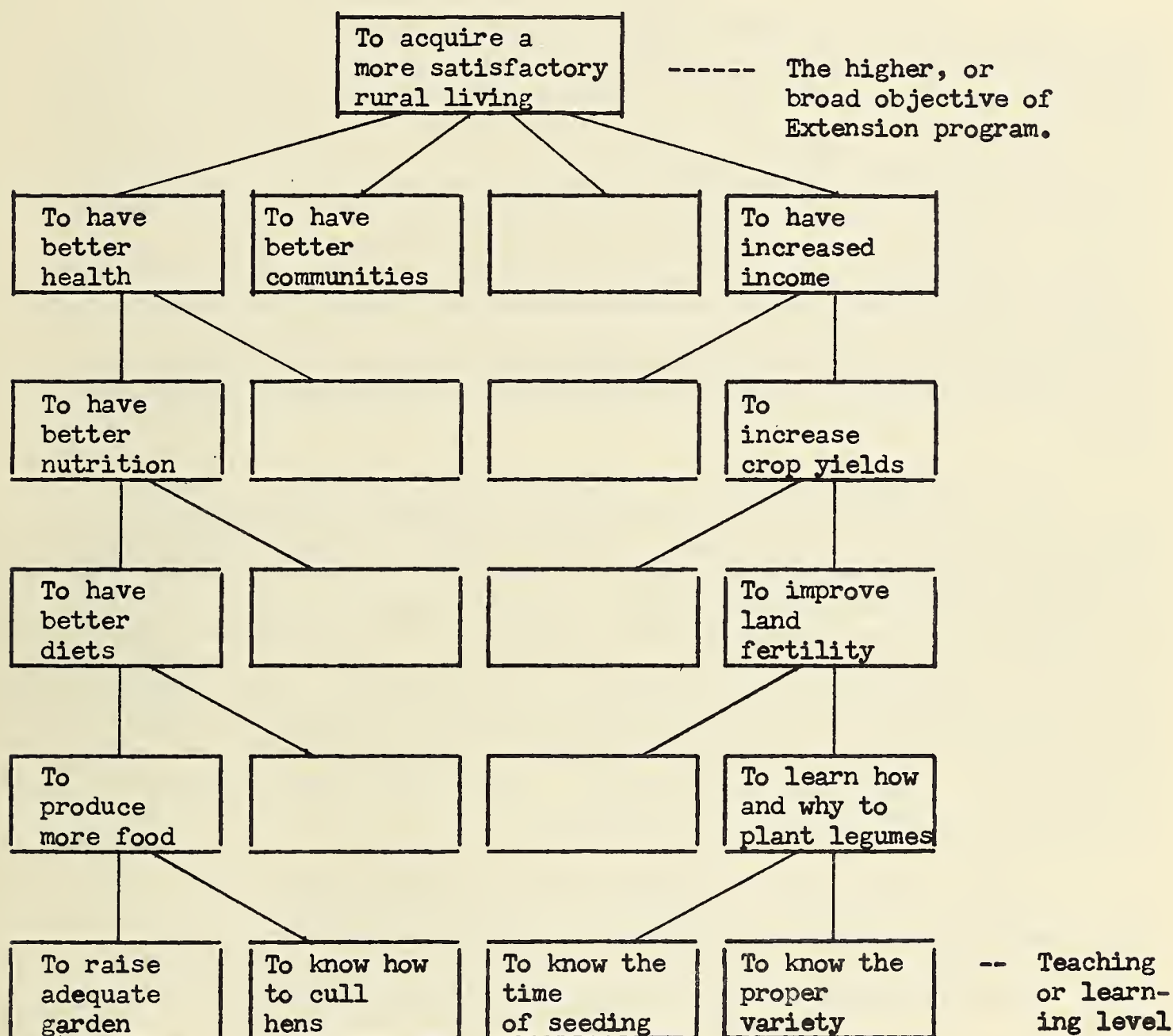
- Soil conservation vs. other uses of land
- Quality vs. quantity -- foods, clothing, and so forth
- Child care -- discipline vs. free expression
- Contouring vs. other methods of cultivation

Notice that none of these indicate doing something for the people. We could carry one item through all these three classifications, such as pruning. Sometimes a farmer expects the agent to come out and do the pruning for him. The agent, instead, more likely than not, expands the farmer's knowledge by giving him information on what pruning is -- he shows him how to prune -- and he attempts to change the farmer's attitude by giving him the advantages of pruning. The farmer should end up by being able to prune his own trees, to decide what trees need it and when, and to apply the knowledge to other trees.

In discussing this classification of objectives, we seem to have diverged quite a ways from some of the statements of the broad objectives of extension work, such as Director Ramsower's "The Ultimate Objective of Extension Teaching is to Promote the Physical, Mental, Spiritual, and Social Growth of the Individual Farmer, His Wife, and His Children." Let's see this transition, and see the very logic back of the way extension work is done -- the consistency of a system which sets up physical, mental, spiritual, and social growth of the people as the ultimate objective and sends the home demonstration agent out to give a canning demonstration in order to reach that ultimate objective.

Levels of Objectives:

The easiest way to think of objectives, and the most usable way to think of them, is in the idea of levels. Objectives have different levels, starting from the top or over-all level, of better rural people, down to the teaching level which involves the carrying out of one small phase of subject matter. I have heard extension workers express their frustration at having to devote their time to such "little" things, when the goal of Extension is such a big, beautiful ideal. There should be no frustration when the worker knows his objectives and knows the weaving of each of his immediate objectives into that big, beautiful ideal way ahead of him.



Each level represents things an Extension worker does to contribute to the objective above it. In setting up a program in your State, county, or community which will contribute to that highest of all objectives, several things have to be kept in mind:

(1) That the level of objective at which you do your teaching is the level of the people. If you have to go way back and teach the farmer which variety of legume he should be planting on his land, then start there. You may have to go farther back and teach him to sow legumes. The teaching level must be where the people are! I believe that you can teach better diets from now on, and your people will acquire knowledge, but they will not go in that direction of better health unless they have more knowledge of how to provide those diets to their families.

(2) The people should realize that they are traveling along the road which leads to one of the higher objectives. There are farmers who have been helped by Extension agents to make more money -- and they have made more money. They reached the objective of increased income -- but have never used that money to bring about a more satisfactory rural living for either their own families or their communities. People, in order to travel the right road to the right objective, have to want to travel that road. I have said that a good objective was one that will provide opportunity for a large number of people to move some distance, and in the direction they themselves want to move. Therefore, in setting up the teaching level of objectives, it must not only be kept in the mind of the teacher, but must also be put in the mind of the people, what the ultimate goal is. Ways and means must be used, if necessary, to change the attitude of the people so that they will want to travel the road all the way.

(3) The teaching level of objective must be simple. It must involve so few points of information, skill, or attitude that people can take them in and get satisfaction. To have better community life as an objective is not a teaching objective. Used as a discussion subject at a meeting, it involves so much that no one point will get across or will be retained. No one point will seem important. If it is broken down to library facilities, or to recreation facilities for the older youth, or to road or electric line construction, it is concrete. The ordinary mind can encompass it, it can be discussed within a reasonable length of time, more people can contribute ideas, it takes on importance, and a satisfying experience is experienced. The teaching must be done in terms of the objectives at this teaching level.

A good test of your Extension teaching program would be to list your teaching objectives -- what you are actually teaching. Work up from that level through broader levels, and see if each does actually and truly lead to one of the broader objectives of Extension Service work. Criticism has been aimed at home demonstration agents spending

time teaching women to paint vases, or make flowers. Justification for the criticism depends on whether or not that activity is leading the women toward a goal recognized by Extension as good and desirable. 4-H Club activities also have been criticized -- as well as agricultural work -- and the criticism is invariably made that they do not seem to be important in terms of the higher-level objectives of extension work.

Evaluation Level of Objectives:

Just as we teach at the "teaching" or working or learning level of objectives, so we evaluate at that level. Our objectives at that level are concrete and clearly defined, and the population which they should concern can be defined clearly. At the higher level of objective, we run into these difficulties in evaluation:

(1) So many other factors also lead to the same goal that it would be almost impossible, if not entirely impossible, to differentiate our success out of the over-all results of all factors (increased income).

(2) Evidences of success are often most difficult to recognize or prove at the higher level (nutrition); the people may not yet have moved along the road of the objectives as high as that level, and therefore there could be no evidences of success that high up. Planting legumes may not lead to immediate jump in income. I can see where soil conservation might lead first to a decrease. In some cases, the teaching level may climb higher and higher as the people advance in their level of knowledge, skill, and attitude. Such a myriad of factors lead to each Extension objective, that I cannot see ahead to where Extension workers will not see new ways that they can help the people move in the right direction.

In clarifying teaching objectives, there are two criteria for them that I want to emphasize:

- (1) They should be defined in terms of behavior or changes in people.
- (2) They must be achievable by the level of maturity of the group of people who are supposed to reach it, and permitted by their available resources.

And there is a rule for writing them that simplifies the clarification of them. A good educational objective has two definite parts:

- (1) The change in behavior on the part of the people that you wish to bring about (change in skill, knowledge, attitude).
- (2) The subject-matter in which this change will be brought about (nutrition, handling animals, etc.).

Discussion III. TEACHING OR OPERATING PROCEDURE AS A BASIS FOR EVALUATION PLAN

Just as definite as the fact that you cannot evaluate, or find out the worth of any activity or program unless you know the objectives of your work, is the fact that you cannot evaluate, or find out worth without a clear picture of the program you followed in helping the people reach those objectives. This is a basic and vital step in the evaluation process. Evaluation is determining how effective your work is; therefore, you must find out if your work contributed to the results.

Methods Related to Objectives:

To take an extreme case, how satisfied would you feel to evaluate in terms of your objectives if not an activity that you carried out could possibly have led the people in the direction you had wished they would go?

In this step, in either planning a teaching procedure or planning an evaluation project, we determine whether the objectives are real or only ideal -- in fact, we determine what actually were the objectives of the program.

In teaching, after clarifying the objectives, the next step in the educational process is planning for and defining the "learning experiences" to be provided for the learners. Essentially, learning takes place through the experiences the learner has; that is, through the reactions he makes to the environment in which he is placed. It does not take place through what you do as a teacher but through what the student goes through as a learner. One teacher -- but as many learning experiences as there are students.

In order to make this planning and analysis more meaningful, we have some principles for selecting or judging the teaching activity (program execution) or carrying out of the job.

You will remember that yesterday I gave you some characteristics of teaching objectives that can help you develop them or judge those already developed: At level of people, achievable by the people, level known to the people, simple, defined in two parts: In terms of changes of behavior of people and in the subject-matter field involved.

Principles in Selecting and Judging Learning Experiences:

1. Set up so people have chance to practice the kind of behavior expected.
2. Such that the people get satisfaction from carrying on the type of behavior.

3. Must be appropriate to the people's level.
4. Many learning experiences can lead to acquiring same educational changes.
5. One learning experience can lead to acquiring more than one educational change.

Your objective may be to help boys and girls in 4-H Clubs acquire an ability to carry out the duties of club officers. That objective certainly can be a most admirable one and valuable one. The purpose of your evaluation might be to determine to what degree the club officers are functioning as they should be. You believe that the club president should know the parliamentary rules, and be able to enforce them politely and diplomatically, be confident and poised, and speak clearly and loudly enough. You believe that the secretary should keep a complete record of the meeting, and read it clearly and loudly enough. And so forth.

Before you can go ahead and start looking for evidence that the officers are functioning, you need to do some desk work. Put down and study the activities you carried out in training the club officers, or in training local leaders to train them. Did you teach them how to do each of these things that you are expecting of them? Did you see to it that the information needed to do these things reached every person who needed it? Was it in terms that they could understand and use? Actually, what knowledge or skill or attitude ought to change in these boys and girls as a result of your activity and teaching?

How you did your work? What did you try to teach? Whom did you try to teach? What did you expect the people to know as a result -- to be able to do?

After you have clarified exactly what your objectives were in your work, what goal you hoped the people would reach, what the people could be expected to learn and acquire as a result of the teaching that you did, and which people could have acquired the knowledge, you have the background for an evaluation project. After these are all made clear, the other steps in an evaluation plan are much easier.

I think you can realize now that evaluation includes not only a measurement of the outcome of your work, but also an analysis of the activity leading toward the results. This analysis ought to take place during program planning and execution, but if it has not been done then, it must be done during evaluation.

Discussion IV. INFORMATION TO BE COLLECTED.

After you have selected and defined the objective for evaluation and the work you did to reach it, the next step in the development of your evaluation outline is to determine what evidence you will need to look for so you will know that your objective has been reached or that you are going in the right direction. This is true regardless of the evaluation procedure you will follow -- whether you will use a record form, make a home visit to determine the progress made by a cooperator or demonstrator, or will carry out a survey of the people whom you have been trying to reach.

Levels of Evaluation:

1. In terms of changes in behavior of people. One level at which to measure your progress or attainment as an educator, teacher, or organizer is to find out what changes have taken place.
 - a. In the people themselves. Have they -- the farmers, homemakers, or young people changed their attitudes or skills, or done anything as a result of the Extension activity or method?
 - b. In the people you are teaching or training, so as to help the people learn something new -- county agents, local leaders.
2. In terms of opportunity. Sometimes it is difficult or impossible to measure progress at the level of the changes in people themselves. In these cases it is desirable to measure our work in terms of learning situations we have set up for the people to learn and to change behavior. If no written materials go out on a given subject, no talks are given, no demonstrations put on, no visits made; then one cannot expect the people to have learned anything as a result of Extension work. Also, it is logical to assume that meetings which have 1,000 people in attendance have a greater chance of changing behavior of a lot of people than no meetings at all, or meetings with an attendance of 100.

The items listed on the first two pages of the annual statistical report form, ES-21, provide a method of measuring the work of Extension agents in terms of opportunity. The more or the better the opportunity provided, the better the chance of accomplishment. A well-organized and well-attended 4-H Club event should offer better opportunity for boys and girls to develop than a poorly organized and poorly attended event.

Therefore, to measure the goodness of organization, of attendance, of coverage, of knowledge -- gives you some information about your success. But it still does not in any way measure quality of educational change in behavior of the people who were organized, who attended, and so forth.

Type of Behavior to Measure:

Because we often measure at the opportunity and not at the educational-change level, we must consider both in deciding what type of behavior to measure.

How can you tell the status or that you have made progress or attained the goal toward which your objective was aimed? What is the evidence of the status or progress? Education is successful when it has caused a change in the "right" direction. Evidence does not always seem obvious. Unfortunately, the human being is a complicated creature, and changes occurring in him are not always obvious to the naked eye. The objective chosen for evaluation as we have seen, must be in terms of behavior, educational or otherwise.

Behavior, when considered as the result of education, is often not in terms of an immediate physical action. It may be in terms of an improvement in skills, a better understanding of a concept, an increased ability to solve problems, a changed attitude, an appreciation of different things, a shifting of values, a change of interests, an adoption of improved practices, an increase in knowledge.

To select the types of behavior you will measure, a careful check must first be made of the types of changes of behavior you have been or will be trying to bring about. This involves studying carefully your background information. As you know, different teaching methods or techniques often have as their objectives different types of changes of behavior. The important thing is to ask yourself constantly: Was this change to happen as a result of my having this particular objective in my teaching plan? Could my methods succeed?

A radio talk may be aimed at producing a desired action or at increasing the knowledge of the listeners. A recipe in a newspaper may be put in for the purpose of encouraging people to use food more economically, tastefully, or nutritionally; the radio talk, news article, or project lesson may be aimed at teaching why this is important. One attempts to bring about a change in practice; the other to bring about a change of knowledge, attitude, or appreciation. We shall need to decide which types of behavior we want to measure -- only part of them, or all of them.

Levels of Behavior at Which to Measure:

Our next decision to make is: Do we need to establish a base from which to measure progress? Measurement may take place at three levels of attainment. First, before any change or any further change occurs -- the people's behavior before Extension has done any teaching or has done further teaching. This place or level where the people start from in their change of behavior is sometimes called the benchmark. The decision whether or not to establish a benchmark has to be made at this step in

the evaluation outline. If you want to know what progress has been made, to have some idea of the benchmark is essential. If you know pretty well what it is, there is no need to put forth effort to establish it. If you are mainly interested in whether the goal is reached and not in how far people progressed to reach that goal, it is not necessary. (Ex.: Status Survey)

A second level would be at any step in the progress of the people toward the ultimate goal of the objective. In evaluating a long-time objective, progress reports are necessary to tell you if the people are making any progress -- to tell you if the teaching methods you are using are effective -- to tell you which teaching methods are most effective -- to tell you which steps have been accomplished and which ones must be emphasized in the program of work.

(Bring in discriminating evidence -- conservation and increased income.)

The third level for evaluation would be to measure the attainment of the final goal of the objective. When you find out if few, some, or all the people have reached that goal, you know whether to retain the objective in your plan of work or to substitute another one: You can find out whether certain teaching methods have been effective or not, and under what conditions.

Two of these levels of evaluation may be incorporated into one testing program or device. Because it is often necessary to establish a base for measurement, very often the plan calls for two levels -- the benchmark level together with the progress -- or the attainment-of-goal level.

Steps in Determining Which Evidence of Behavior to Look For:

Our problem, then, is to decide which changes in people are to be considered as evidences of status, progress, or attainment. There are several steps to follow in selecting these evidences.

(1) Decide whether the things done -- the activities carried out -- or the things we can count, will be adequate as evidence. These are the easier results to use in evaluation, and they make much better working material for evaluation during the early days of your evaluation experience. If, however, these types of evidences are not adequate proof of the status, progress, or your having reached your goal, then it is necessary for you to think in terms of intangible evidence, such as changes in knowledge, attitude, or thinking.

(2) An important factor in deciding which behavior you will measure is your own facility in measuring different types. There is no question about the fact that it is easier to measure changes of behavior which have tangible evidences for proof than it is to measure those which

have intangible evidences. The fact that a woman has made use of a recipe she has seen in a paper is easier to determine than the fact that she now better understands the value of good nutrition in child development. The fact that a family gathers around the table to plan the duties of each family member is not difficult to ascertain. But whether the family now gets along better as a result of this planning around the table is an intangible result that is hard to get at.

To determine the real results of extension teaching, we cannot ignore these intangible results just because they are hard to evaluate. On the other hand, when one is becoming accustomed to the procedure of evaluation, it is highly desirable that the techniques, methods, procedures of evaluation be learned while measuring those results which are easier to measure -- such as practices adopted, or perhaps knowledge gained. When the techniques are mastered, they can be applied to that type of measurement which is more difficult. Just as we learn to read by starting with simple words, short sentences, and simple stories, so in the same way we learn evaluation methods by starting with the simple, and advancing to the more difficult. We just want to be sure, if you do begin with the easier methods, that you do not feel that all evaluation is concerned only with tangible results and results that we can see and count.

(3) Limiting number of evidences -- Decide which of these evidences you will look for in your evaluation plan. If your list of evidences is very brief, and to look for them all would not require too much work, include them in your plan. If your objective for study has been defined and limited so that it requires only one, or a very few actions or changes in behavior to reach it, only one or very few evidences need be looked for.

However, there are more often numerous evidences of change of behavior. When such is the case, there are three ways of selecting which ones you will look for in your measurement:

(a) Study your list of evidences to see if evidences of some changes; if discerned, would be clues to the fact that other changes must have been brought about. Sometimes only these discriminating evidences, or those that are clues to others, are enough to look for. In a progress report, however, be careful that those actions, to which others point, are not omitted, in case they indicate steps in progress.

(b) If there are no discriminating evidences, or if there are too many of them, study them over to see which indicate the most important changes in behavior. If you can't look for all changes, certainly those you consider most important will be of most value to you.

(c) If your long list cannot be reduced by either of the above methods, your list can be reduced by choosing at random a practical number of them. Random methods of selection will be explained fully in the discussion on Sampling.

(4) The time factor must be considered -- Seldom is a change of behavior, whatever the type, apparent immediately upon reception of the teaching. For types of changes such as in attitude, appreciation, knowledge, or values, the change may take place immediately. Even in many of these cases, however, the respondent needs time to think over the material, weight it in his mind, and be exposed again to his customary sources of information, before a change that has any permanency actually takes place. Knowledge which is evidenced immediately after a presentation of the material may be forgotten by the next day.

When the change of behavior desired is a change of practice, the time element is extremely important. First, some things cannot be done until, for example, a certain season has arrived, an opportunity has arisen to get certain materials, money is available to buy what is needed, or the family is all together. Some things take a long time to do, such as remodeling a house or raising vegetables. Some things take place in steps, as planting a garden, harvesting it, canning the food, and serving balanced meals all during the winter.

In planning the study, therefore, it must be decided when you can reasonably try to find out if a certain practice has been adopted. Judgment must be used in this phase. We don't know exactly how long after a person has had a chance to carry out a practice, and does not do so -- we should wait to see if he is ever going to do it. The longer we wait, certainly the less likely it is that he will ever carry it out.

But, on the other hand, there is a limit to the time we can wait and still expect to find out what was done in the past. For daily happenings, 2 or 3 days can break up the memory span. What did you have for dinner 3 days ago? For weekly happenings, 2 or 3 weeks will do the same damage. For seasonal activities, if the activity was important in his life, one can expect a respondent to remember during the rest of the year. Most homemakers remember how many quarts of food they canned for most of the year, but if canning has started for this year and you ask how many quarts she canned last year, the accuracy of her memory has already deteriorated.

When reporting memory items, it should be explained that the data are from memory and not from records, as memory does distort data -- increasing them if a large number is desirable, decreasing them if a small number is desirable. For some types of information, you will find you will need to collect data at two different times, because of the time element entering differently into the different actions or behaviors desired.

In what units or terms will evidences be summarized? Will they, for example, be by degrees of attainment, amounts of things accomplished, or length of time an action has been carried out? Will your evidence of a year-round garden be in terms of certain vegetables raised, in amounts of certain vegetables canned and stored, or in other terms? Will it be the

average number of quarts of vegetables per person, or will it be the percentage or proportion of the people who canned any determined average, or a minimum amount? Which will be the evidence you want, or will need, to form a basis for your program planning? Will direction of change of attitude toward a social problem be adequate, or will you need to know how far that attitude has changed -- and what resulted from the change?

Situation in which measurement can be carried out:

After deciding what evidences you are going to look for, in order to evaluate your progress or the need for a program, you will need to look for situations in which you can find this evidence -- situations also which lend themselves to measurement.

People must have a chance to display the change of behavior you are trying to bring about; they must also have had some motivation to make a change. (Lafourche women). A woman might have the money, material, time, and energy to refinish or repair her furniture; if she has new furniture and plenty of money and labor to have any repairs done for her, and she has no motivation to carry out recommendations dealing with dairy herds; a mother of grown children can't be expected to make a change in practice dealing with feeding of preschool children. So we must decide where we will find the people whom we are trying to help, and find them in situations where they can change because of education.

(5) Descriptive information or face data needed: Face data are descriptive information you find you need to obtain from or about the same specific individuals from whom you get your evidence. You ask yourself what other information you need about the people whom you are teaching. What other information will you need to help you interpret your findings? If we find out a method or a project was successful, do we not need to know with what kind of people it was effective? If it was ineffective, what factors may have contributed to its ineffectiveness?

Think over the teaching methods used, the information material given to the people. If you have been expecting the people to have any certain mental or material possessions or surroundings in order to make a change, you should know if the people have them. Do you expect certain types of people to react more than others to your teaching? What would you like to know about the people who made a change in behavior and about those who failed to react? Naturally you seek information which will help you -- especially after you have ascertained the differences between the groups -- in making new plans for the next year's work. Much information, of course, about people is available in county records and Census materials. But beyond that, what else are you going to need to help you the most in your future work?

We are now ready to go into a more detailed discussion of another point that I have mentioned we would refer to over and over again -- the people, or the population. The other points were the what, or objectives, and the how, or methods used. Now we are at the "whom." From whom shall we gather the information which will tell us of our success? We have gone into this to some degree in connection with the discussion of your teaching plan. Now we must make a final decision on this, as it is impossible to perfect your method of getting the information on success without definitely knowing from whom the information will be collected. And the decision about the whom is determined by the what and how. And the decision about the whom determines how you will gather the information.

In your program planning, you decide whom you would like to reach or teach. Then, by the methods you use, you define or limit those you do reach or teach. Meetings are a method that reaches only those who attend meetings; therefore, when you teach something only at a meeting, you would gather proof of the success of that teaching from those who attended the meeting. Information given over a radio program reaches those with radios who listen to the program; information taught to local leaders reaches those whom the leader teaches -- a more difficult group to define.

These people that you reach, and you cannot help them learn until you reach them, are called in more technical terms your "population." In planning your work, you should define your population; in evaluating your work, if you did not define it during your planning, you need to define it then. To whom are you interested in getting across the information or from whom are you interested in gathering information? Everyone in the county or State? Only the farmers? Only the adults? Only those doing dairying? Only the women attending meetings? Only the people receiving bulletins? Only the people on your mailing list?

The "what" and "how" determine the "population" which determines the method used to gather information.

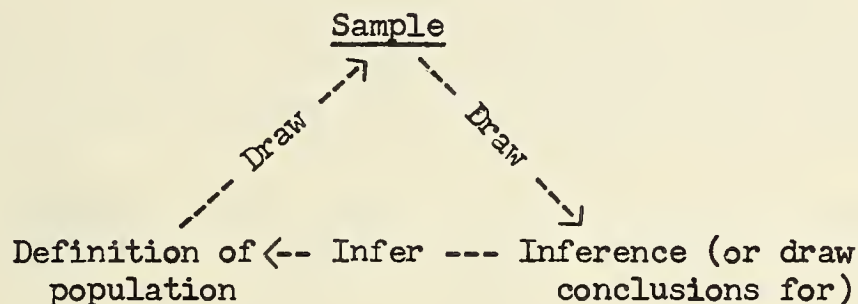
Extension agents reach and teach more people than they can possibly get to to determine whether the people have experienced an educational change. In a way this is unfortunate, as, in most cases, the way to get complete and reliable information is to go to all the people you reach and find out from them in one way or another whether they learned each of the different things you were trying to get across. Since Extension workers cannot do this, it becomes necessary to get your proof of success from only a part of the people about only part of the things taught.

When you get evidence of success from only part of the people, you are "sampling" your "population." Since the method you use in selecting this sample determines so greatly the reliability of the evidence, I am going to go into it in some detail. Even though none of you, or very few of you, will ever develop a very comprehensive sampling plan yourselves, the philosophy back of correct sampling seems to be best explained by discussing the practical aspects of actual sampling procedures. The ideas followed in such procedures should be in the mind of every teacher or organizer when he is evaluating his own success.

General Definitions and Principles of Sampling:

I think it will be best, to begin with, that we first are clear as to what we mean by sampling, or at least what I mean by sampling. Sampling means selecting, from the whole group concerned, a small number, from which an estimate can then be made for the whole group.

"Sample" as most of us have thought of it, is defined by the first part of this definition -- "Sampling means selecting, from the whole group concerned, a small number." "Sampling" as used in statistics and in any evaluation work must have that last clause added -- "from which an estimate can then be made for the whole group." Sampling involves four phases:



This definition brings out three steps involved in sampling, which must be considered.

The size of the population decides whether a sample need be taken. If the whole population can be reached with a practical and reasonable amount of time and expense, there is no need to go only to a sample, and as a result reduce to a certain extent the reliability and accuracy of your data.

The second factor is concerned with the words "from which an estimate can be made for the whole group." This indicates right away that the sample from whom you get the information must be representative of the whole population -- it must represent all the characteristics found in the whole population, or else it cannot be used to make a good estimate for the whole group. The answer or results one gets from a sample should be a close approximation to what one would get from the whole population.

Selection of a sample for your evaluation determines to a large degree the reliability of your evaluation. A device or technique is reliable if the result you get is a close approximation of what the true result is. Out in your county there are definitely a certain number of people, or a certain percentage of the people, who actually have changed their housekeeping or farm practices as a result of Extension teaching. If your evaluation of your teaching of such practices is reliable, the percentage of people that you find changing practices should be very near to the true percentage.

Two things are of interest to you and of importance in selecting a sample: Size and representativeness, both in relation to accuracy.

Size:

One of the common errors in the selection of samples is to take refuge in the size of the sample, as though sample size by itself could give assurance of representativeness. The size of a sample is no proof of its goodness or representativeness. If the sample is not representative, size doesn't help any. (Literary Digest poll.) A sample of 1,000,000 improperly selected can be less accurate than 100 properly selected.

If a method of sampling is planned which should select a representative sample, then the larger the sample is, up to a certain amount, the more accurate will be the results. If such a procedure can be followed, add cases until these additions fail to make any important differences in results.

Example:	<u>Cases</u>	<u>For</u>	<u>Against</u>	<u>No opinion</u>
	20	50%	45%	5%
	50	20	70	10
	100	32	61	7
	150	31	62	7

As long as the whole "population" is not studied -- a sample of the "population" is selected -- the results have what is called an "error," or "standard error." This does not imply a mistake; it is a measure of the probable difference between any result you get and the true result. The larger the sample you use, the smaller becomes this "standard error." In most studies, you are not as concerned with the exact number or percentage of people who have made some certain change in behavior as you are in whether or not a large percentage changed practices, or twice as many perform a certain way after the teaching as before. However, there are statistical methods of determining how large your sample should be to keep the "standard error" down to a certain size.

Common sense should be used in final determination of size of sample.

The size of the population, in itself, has nothing to do with the size of the sample. It is the size of the sample which is important. A sample does not have to be a fixed percentage of the "population." Size depends on desired accuracy, cost, homogeneity of population, and desired breakdown of data for analysis.

The cost factor includes time and labor as well as cash outlay. Since the cost factor forces us always to resort to the smallest sample we can use and still retain a usable margin of statistical error, we first of all have to put a great deal of time and attention to making that sample the best it can possibly be as to its representativeness of the population. As to its size, it must be large enough to include all the different influencing factors. If age, education, farming status, nationality, distance from town, and other factors affect the response of rural people to Extension influence, then the sample must be large enough to include people of all classes within each of these factors. That is, people of all ages within the age factor, of different education levels, farm owners and farm renters, American born and foreign born, near to town and far from town. Then, not only must the sample be large enough to include all these classes within each factor, but also, if any analysis is to be made of any one of these classes, such as people with less than 8th grade education, enough records must be obtained for that class to enable a person to study it. For example, if the number of women 20 to 25 is very limited in home demonstration clubs, but you do want to know that programs reach them, or which practices are most effective with them, then the whole sample must be large enough to include an adequate number of these women of this age group. Personally, I don't like to study a class with less than 30 records in it -- I'd prefer to have 100. If a class is to be broken down again within itself, as dividing women 20 to 25 into education classes, it has to be larger, of course. The number to be included in a sample has to be determined each time a sample is drawn. The more heterogeneous a population is, the larger must be the sample; the more homogeneous it is, the smaller it may be.

Sample size is determined by:

- (1) Size of sample desired for statistical accuracy.
- (2) Heterogeneity of population.
- (3) Proportion in population of lowest represented characteristic you want to study.
- (4) Cost.

Distortion of data may be introduced in two ways: First, by incorrect selection of the sample -- either because the sample was not drawn from the true population, or because the method of selection was wrong; and second, because of those in the sample who did not return the records or fill them in (mail questionnaires!). Another bias can creep in through poor questionnaires or badly asked questions.

The use of sampling implies, then, the obtaining of information from a small number of carefully selected people whose answers will be typical of a larger number of people. There is no need for random sampling if there is no intent to apply to a larger group.

Methods of Drawing the Sample:

Random selection is one way of getting a representative sample. Random sampling means selecting each item or person in such a way that each item or person in the universe to be studied will have an equal chance to be selected. A sample chosen by random selection will more likely have the different characteristics in the same proportion that they exist in the universe than in any one other proportion one may name. A sample of adequate size will more likely than not resemble the universe from which it is selected.

I have used the word "selection" in connection with random. This is because random sampling calls for careful application of rigorous techniques. Random sampling is not just a "hit or miss" method of picking out items or people. It is not enough to know that certain biases have not been introduced at the point of sample selection; it is also necessary to know that all parts of the universe are represented.

Several methods of drawing samples have been used in Extension evaluation work, especially in field studies.

List sampling: -- One method of sampling is to select a random group of people from a list. Random selection means that every person on the list has just as much chance of being selected as another. The list must, of course, comprise the total population in whom you are interested. This method can be used when you do have a list of the total population. An illustration of the use of this method is the Chemung County study in New York State. When a study of the county situation revealed that every farm-woman-Home-Bureau member (farm being determined by their own definition) could be interviewed, but that there were more farm women who were not members of the Home Bureau than it would be convenient to interview, a list was made up of all the farms in the county. The farms on which a Home Bureau member lived were deleted from the list. The other farms were grouped according to the community in which they were, and numbered within each community. Then a table of random numbers was used to decide which of these farms to go to. In this case of Chemung County, enough farms were selected from each community to make each community represented in the sample in the same proportion as it was in the total population.

I want to call your attention here to what the population was for this study. It was all the farm homemakers in Chemung County on farms which qualified for the definition of farming used by the study committee. Therefore, in using the data secured from this study, we cannot say that they are true for all farm people in Chemung County classified according to the Census definition, or for all rural people, or for all people in

Chemung County, but for all farm women living on a bonafide farm in that county. The data from the non-Home Bureau members will be used as an estimate of what is true for all non-Home-Bureau bonafide farm women in Chemung County.

In a study of the Connecticut Homemaker in Connecticut, the list of random numbers was used in a different way. In each county there was a mailing list of the women receiving the Homemaker. It was decided that one-twentieth of all the women receiving the publication would give a large enough sample. Then, within each county, the towns were listed alphabetically and within the towns the women were so listed. A number was taken "at random" from the table of random numbers to decide which name to start with in the county. Then, starting with that name, every twentieth woman was selected.

With this random method of selection, the results obtained from the sample can be considered a reliable estimate of what they would have been had they been obtained from all women taking the Homemaker.

Area sampling:-- Because efficiency per dollar is an important factor, there have been ways of sampling devised which save travel and time with possible increase in statistical error. One of the most frequently used is area sampling. This means that one takes a random sample of areas in which people live, instead of a random sample of the individuals themselves. In this way, instead of the people probably being scattered all over the county (or State), they will be grouped together so one interviewer can carry out several interviews in one section with very little driving between interviews.

The BAE Master Sample is the best known use of this method of sampling. They have sampled the farming areas of the whole United States, and the sample includes one-eighteenth of all farms. To do this, they marked off on a county map for every county in the United States small areas containing about an average of five farms, bounded when possible by observable boundaries, like roads, rivers, railroads, and so forth. Then, within each county, every 18th area was selected, and if the sample were used, every one of the five farmers living within each eighteenth area would be interviewed. Although this method decreases the accuracy of the estimate made from the same amount of data, more data can be collected for the same expenditure of time and effort.

A method similar to this one was used in Massachusetts in a study of home demonstration organization. The population was decided to be women living in rural areas in the six counties cooperating in the study. Degree of organization varies from town to town within the counties. Three degrees of organization were agreed upon and the towns were classified according to these degrees of organization. The towns of any one degree of organization were grouped together for sampling. Areas bounded by distinguishable boundaries were delineated on maps, each area having an average of eight homes to an area, and every 11th area was chosen for

the sample. Every home within these areas was interviewed. In some areas everyone had moved out; in others as many as 30 new homes had sprung up; but by interviewing every home, the sample remained representative of the towns, as this process of people moving out or in was taking place everywhere.

In the recent Vermont study of Extension work, a similar plan of area sampling was carried out over the whole State. For an Extension Farm Labor study carried out in Illinois, the BAE Master-Sample material was used -- the statistical laboratory at Ames, Iowa, drew a sample of the size the Illinois Farm Labor Office wanted from their Master-Sample maps. For drawing this sample, the Illinois Farm Labor Office was required to define carefully their population -- such as, what did they want to call farms, did they want farms in incorporated areas as well as in open country, did their population include all farms or only those hiring labor?

In Ward County, N. Dak., where area sampling was used, advantage was taken of the fact that out there, every section or square mile of land is bounded by a road and 36 such sections make up every township. The southern half of the townships were taken as their sampling units, and these were chosen at random over the county. With farms being quite consistent in size in any one county in North Dakota, we would know to begin with about how many farm homes each such half-township would include (perhaps about 16 in that area).

Grid or random point sampling. -- Another method of sampling geographic areas is to use a "spot" method. One way is to lay a grid or screen over a map and choose the points where the wires or lines cross. Another way to choose the "points" is to put a number of dots, hit or miss, over a piece of paper and lay it over the map.

Then, either all the homes within a given circle-radius are included in the sample, or an "n" number of homes closest to the spot are chosen.

Both of these methods are faulty if the distribution of homes is uneven.

Group sampling. -- The sampling of groups of people instead of individuals follows the same principles as does area sampling. We often sample home demonstration clubs or 4-H Clubs instead of only sampling individuals within all club memberships. Doing this increases statistical error, and decreases cost. If heterogeneity is as great within groups as between them -- then we can take big groups.

A variation of group sampling involving double-stage sampling may produce the most desirable accuracy-cost ratio. This means first choosing at random a certain number of groups or clubs, and then sub-sampling from these clubs, selecting at random only a certain percentage of the members belonging to the selected groups or clubs.

As with area sampling, 5 individuals chosen at random from 20 groups chosen at random (giving a sample of 100) costs more but is more efficient than 10 individuals chosen at random from 10 groups chosen at random (also giving a sample of 100).

Stratified sampling. -- These methods just described illustrate methods which do not aim at bringing into the sample any certain proportion of people with certain characteristics. With the methods I have mentioned, they will by random selection be represented in the sample in the same proportion that they are represented in the whole population.

Sometimes, when we know to what degree certain classes of people are represented in the population, we take advantage of this knowledge, and make an attempt to get that definite proportion of certain characteristics in the sample. This method was used in two counties in Ohio and in Waldo County, Maine. In these counties, all land was classed by land-class, and we knew about to what degree each land class was represented in the county. Feeling sure that land-class on which people lived was a very good index of the socio-economic status of the people, we sampled land-classes, not people. In Waldo County, we found one area in the county in which all land classes were included in one solid area. We interviewed every home in this area, got all land classes represented, and found when our data were studied that they were representative of the county, as a whole, as we had suspected.

However, there are times when it is desirable to include a greater proportion of some one class than exists in the actual population. In these cases, deliberate effort is made to get more of one class or another, so that enough are included in the sample to study. When Miss Collings sampled all home demonstration agents in her time study, if she had included the same proportion of Western-State home demonstration agents in her sample as there were in the population, she wouldn't have had enough of them to study -- the sample would have been predominantly Southern home demonstration agents. Therefore, she took about an equal number of home demonstration agents from each Extension region, and analyzed them by regions. When making a United States total, in order to make her calculations a good estimation of what actually occurred among all home demonstration agents in the United States, she had to weight the data from the Central and Southern States much more than those from the Eastern or Western States, and the weighting of the Western States was very much the lightest of all.

Quota sampling. -- Quota sampling is a refined variation of stratified sampling. In quota sampling, an attempt is made to include in the sample some of every known kind of individual in the total population. If possible, the same proportion of each is included in the sample as exists in the population. If this cannot be accomplished, weighting is used in calculating totals.

This method presupposes very complete statistical information about the total population, as well as enough basic information, concerning which of the known characteristics would influence the value of the unknown characteristics, to determine which need to be sampled.

Judgment sampling. -- The sampling methods described above depend on objective procedures which should result in representative samples -- procedures which give everyone in the population an equal chance of being selected. There have been samples, however, chosen on the basis of the judgment of the selector. In doing so, the sampler hand-picks each individual to go into the sample, making up what is intended to be a sample representative of the population from which chosen.

Yates ^{1/} has this to say about this method:

"The ideal which is aimed at in sampling is to make the sample as representative as possible, so that measurements or observations on it can be taken as virtually equivalent to similar measurements on the whole population. The fact that this ideal is in the mind of the sampler when taking the sample naturally influences his selection if he has any freedom of choice. Most samplers when selecting a representative sample will deliberately reject abnormal material, or if they feel that the sample should be representative of the abnormal as well as the normal will deliberately balance up the different categories abnormally.

"Unfortunately the sampler's claims to be able to select 'a representative sample' by personal judgment are largely unfounded, and his selection is in fact subject to all sorts of biases, psychological and physical. To avoid these biases and to provide an estimate of the representativeness of the sample, i.e., of the 'sampling error,' more rigorous processes of selection have been devised."

Houseman ^{2/} in Agricultural Economics Research, comments, "It is reasonable to expect that judgment samples will usually be selective in the direction of what the mind regards as the more important elements or in the direction of the less obscure elements. It is also reasonable to expect in general that the variability would be less within judgment samples than within the whole population; if so, this would mean that distributions based on judgment samples might be seriously distorted. ... Judgment samples naturally look good to the person who made the selection, but since a sample cannot be appraised by its looks, judgment samples are likely to have erroneous impressions of the accuracy of their samples."

^{1/} Yates, F. Some Examples of Biased Sampling. Ann. Eugenics 6:202-213, 1935.

^{2/} Houseman, Earl. Designs of samples for Survey. Agr. Econ. Res., vol. 1, No. 1., January 1949. (Published by the Bureau of Agricultural Economics, U. S. Department of Agriculture.)

Limiting the Population to Study Size. -- Within certain populations, especially when variability is great, it is often found advisable to so specifically define the population that all or nearly all of the members of the population which have definite, described characteristics, are included in the survey. In such cases, sampling, as defined at the beginning of this chapter, is not necessary.

This type of sampling presupposes a good knowledge of the population -- and most important of all, means that the analysis of the data must always be in terms of those characteristics. It is often a dangerous method, inasmuch as unknown characteristics may enter the picture and cloud the data. This method has been used several times in 4-H Club studies. In our study of High-School Youth and 4-H Club Work in the New England States, we very carefully selected clubs which had been in existence 2 of the last 3 years, clubs which had had the same leader 2 of the last 3 years, clubs which had on their enrollment boys and girls who had dropped out when 8th graders or Freshmen, or had present members now Freshmen or Sophomores, and clubs in rural or suburban areas (not urban). By the time these qualifications were applied to the clubs in the counties included, every club with those characteristics in five of the six counties was included. In only one county were there more than enough such clubs. In that county we listed the clubs alphabetically and went down the list until enough had been included. As a result of this narrowing of the population, we continually have to remind the listener, or the reader that the findings apply to only those clubs and their members.

Testing the Sample:

We find that in the informal evaluation or study work that we do, we usually do not do much formal testing of the sample. This makes it vital, of course, that a particularly careful job be done in planning the sample and taking the records.

If the "universe" to be studied is a class of people for whom information is available from the Census, we can check data from the study against Census data. If information such as is requested for the Census is not considered important for our evaluation itself, we find it necessary to ask some extra questions for the lone purpose of having data to check with the Census.

Another test for adequacy is to draw from the "universe" several other groups of approximately the same size as the given sample. Then the results for the different groups can be compared. Dividing your sample in two parts and comparing the results of the two is another way.

Discussion VII. HOW YOU WILL GET THE INFORMATION.

There are many ways in which an Extension worker can gather evidence of the success of any activity, whether educational or organizational. However, they all have two characteristics in common -- they must be so carried out that they measure the success in terms of the objectives of the activity, and the information must be collected from a representative sample of the total population which the worker attempted to reach.

Each way of collecting evidence or information has its advantages and disadvantages or shortcomings, and these should be taken into consideration in making decisions as to which to use.

The following ways come to my mind immediately:

1. Mailing a questionnaire to the people.
2. Having leaders ask their neighbors or members of their groups or clubs.
3. Distributing questionnaires at meetings.
4. Asking, orally, questions in meetings and asking for a show of hands.
5. Asking office callers, or listening to their reports.
6. Asking people during farm visits, or observing during farm visits.
7. Observing results in fields, farm yards, and homes while driving around county.
8. Observing in meetings or events.
9. Getting a complete story from one or a few individuals or cases.
10. Counting requests for materials after radio talk, meeting, or news article.
11. Report forms from organized groups as 4-H Clubs, home demonstration clubs, DHIA, etc.
12. Farm-and-home records from cooperators.
13. 4-H Club project records.
14. Field survey.

Each of these methods, and others, can be analyzed to see whether or not they can provide the Extension worker with reliable and valid information. Let's take the first one first: Mailing a questionnaire to the people.

1. Can the questions on the mail questionnaire be fitted to the objectives of the project in question?
2. What about the people returning them? Are they representative of the "population"?
3. What are other advantages (cost) and disadvantages (misunderstanding of questions) of this method?

Discussion VIII. FORMULATION OF QUESTIONS.

The questions used for collection of data are most vital in determining the validity and reliability of the information you collect. We have been emphasizing that the evidences to look for as proof of success or failure must be in terms of the objectives of the job, and collected from the population you were trying to reach.

The questions asked in a questionnaire or data form, should be so set up that information or data collected separately from a number of people can be added together or summarized to give a total which will give you the picture of the group. There are different kinds of forms:

(1) The questionnaire is made up of direct questions which the respondent answers. There are two kinds of questionnaires - the mail questionnaire (sent by mail to the respondents who return it by mail) and personal-interview questionnaires (handed directly to the respondents, individually or in groups, who hand it back directly).

(2) Another kind is a report form which has boxes or blanks on it for recording the number of any sort of thing, the description of which is included on the report form. The annual statistical report is an example of a report form, as well as your own county and State records.

(3) - There are also questions carried in one's mind while talking to people or observing.

At this stage, the right or wrong questions or record form can still keep your data from being reliable or valid. You ask questions, the answers to which tell you whether you were successful or a failure. These questions must, of course, be so asked that they get the kind of information you want.

What goes into the questions is determined directly from that list of evidences you prepared in Step V. Consider such evidence and then put down the questions you will need to ask in order to get the evidence of success or failure.

There are five criteria which any question such as on a value scale interest check, questionnaire, test, or record form, or in one's mind must meet to be fully effective.

A. Validity -- Does the question measure what you think you are measuring?

What we get from a question of any kind should be what we are trying to get, and not a test of the ability of the enumerator to ask

the question correctly, or of the respondent to interpret the instructions. To get validity means returning over and over again to the objective for study to see that each question asked pertains to what we are trying to measure.

B. Reliability -- Is sample large enough so we can draw conclusions from it? Is the question so formed, and the people from whom the information is requested so chosen, that if the same question were given over and over again, it would get the same results? In other words, is the question trustworthy? Will the results you get approach closely the true results? Does it get a true answer? Is the result a close approximation to what the true result is?

Enough samples of information have to be obtained, or enough questions asked, so that we can draw conclusions from them. In a 4-H attitude test, which has been used often in 4-H Club work, the responses to one problem were not considered an adequate sample of all true responses. Five problems were presented, in order to stabilize, if possible, the responses. To test whether we have helped people improve their diets, just the fact that more milk is drunk doesn't give you enough information to tell you whether the final outcome of your objective has been reached.

The sampling of the population, or of the people whom you have tried to teach, has been presented, in detail. This is a very important phase of the study outline, and on it depends both the reliability and validity of the study.

C. Objectivity -- Does the form of the question prevent the idiosyncracies of the person administering it or interpreting it from affecting the responses? If equally competent people ask a question and interpret the answers, and both get closely similar results the question is objective.

The kinds of questions used on questionnaires are usually classified -- objective or subjective -- according to the kind of answer they require. Some require only a YES or NO answer, others a figure as an answer, others require sentences. Others require the respondent to check in the space which indicates their response to the question asked.

Objective tests or questionnaires or record forms are those in which the respondent gives brief answers to definitely stated questions. It is certainly true that the respondents can vary a great deal in their interpretations of the questions -- on the other hand, the questions have been made up with plenty of time to consider their exact form, and often are approved by several people. From the interpreter's viewpoint, objective tests will be summarized the same, no matter who does it. So many checks in a YES column can add up to only one number. There are various forms of objective questions besides the YES-NO type.

The multiple-response or multiple-choice questions require a choice of one answer from among many. In some of the questionnaires which Extension has used are examples of questions which ask that the respondent check as many as apply. They are not easy to construct. The possibilities of answers have to be plausible. When used to show attitude, it is difficult to present alternatives which represent the whole range of opinion, to present alternatives which are mutually exclusive.

Other types of objective questions are true-false or agree-disagree completion questions, matching questions, and identification ones.

The essay type test, or the free-answer question, is the least objective of all types. In order to judge its use, we must consider it from the angle of the respondent and of the interpreter.

Most people are not accustomed to expressing themselves freely, especially when it comes to telling reasons for or against, or beliefs, and so forth. People are not vocal by nature, and especially rural people whose daily duties do not involve practice in expressing themselves. Descriptions and commands are more common to their ordinary conversations. Many of life's responses are not complete essays. Some of the greatest decisions in life, to be determined by reasoning from facts, experiences, and future hopes, are expressed in one word, YES, or NO. In either answering a free-answer question, or taking part in a discussion meeting, many of the most vocal persons know the least about the subject, while those who talk very little don't get across what they know because they are not able to. People can seldom think of everything they could say on a subject at short notice. We have all experienced that after-thought, "Oh, why didn't I think of saying that?" And too often, the exact thing the respondent didn't think of saying is the answer we are trying to get. In one questionnaire we were summarizing in the office, we found that one idea was expressed in 3 out of some 50 answers -- it happened that one idea was something we knew to be true in all 50 cases. In the other 47 cases the respondents either didn't think to include it, or didn't know it was meant to be included. When we think of those 47 missing answers, we wonder how many of the other ideas we summarized would have been expressed by other respondents if they had thought of them.

On the side of the interpreter, it is very difficult to interpret what the respondents meant. Each has his own method of expressing his attitude or viewpoint -- some wordy, some in very few words. In order to summarize these answers into presentable form, they must be catalogued and grouped. In our office we have found many cases where those who do this job have disagreed on what class some answer falls into. The best method is to have the person who is to report the results also do the summary of the free-answer questions so he knows what he put into each class. In a big study, this is impractical.

There certainly is a place for free-answer questions. If ability to express ideas is what you are trying to measure, it is the only type to use. When an issue is new, when we do not have a very good idea of

what classes of answers, what degrees of attitude we are likely to get, the free answer is necessary. In some of our Extension questionnaires in which we now list possible answers to check, these lists were first started from summarization of free answers. We still allow for free answers to be given at the end of lists. Another place we often use free answers is to check on YES and NO or other brief answers. If we wonder if a YES or NO answer is likely to be a guess at what we want, we add a free-answer question, WHY? In these cases, we are not interested in the WHY at all, but use the answer only to determine whether the YES or NO was the respondent's correct answer. We also often use a rather general free-answer question at the end of an interview, to allow those respondents who want to talk, a chance to talk, to get things off their chest, and incidentally, to see if we can't pick up some ideas that we may have overlooked in the formal questions.

D. Practicability -- Is it practical to use and summarize the questions? In all studies or evaluation procedures it is necessary to consider time, money cost, and convenience. Will it take more time than it is worth, will it cost more than you can afford, will it inconvenience too much you or the people whom you are serving? Any evaluation technique takes some time, and we may say it will inconvenience someone or other because of that; therefore, in planning the time schedule on it, the length of the questionnaire, the time involved in summarizing it, the cooperation we will need to ask of the people, the value of the results must continually be balanced against these costs.

E. Simplicity -- This criterion could probably be included in the one above of practicability. The more simple a questionnaire or other device is, the more practical it will be to apply and summarize. Keep the techniques and forms simple for four reasons -- to make them easy to use, to make them easy for the people to understand, to make them easy to summarize, and to minimize time between field work and the report.

Physical Form of Questionnaire:

The following rules apply to all questions, regardless of the type, and will contribute to acquiring validity, objectivity, reliability, practicability, and simplicity.

The questions: Put down, first, what it is that you want to find out. "I want to know if" Then, ask exactly what you want to know. Make them short, clear, concise, as possible. Then go over each question and ask yourself these questions: Is the question related directly to the purpose of the study and the teaching objectives being studied? Does the question contain any clues or cues to the correct response, or the answer you think you want? Don't start a question with "Don't you think?". Is all of the question necessary? Is the point of the question sufficiently clear? Can there be any question about the definition of every word in the question? Are you sure every respondent will think of each word as you are thinking of it? Will every respondent understand each word? If you use words that can have several meanings, define them.

Place the questions in logical order. A person's mind is not accustomed to jumping around from subject to subject, and it prevents accurate thinking and responding if you try to get your respondent to do so. Start the form or interview with questions which are easy to answer and to which you think the respondent will be willing to answer. Questions which require deep thinking right at the beginning will frustrate the respondent. Questions which the respondent will resent, until he understands what you are getting at, should come later in the interview or form.

Provide for definite, understandable answers. Make it clear in what terms you want the answer. Don't say "How often do you serve green vegetables?". Make it, "How many meals a day, or days a week do you serve ...?".

When providing possible answers to check, even if only YES and NO, be sure that you provide every possible answer. A "No opinion" answer must usually be provided for many questions. An opinion cannot and must not be forced. During an interview, the enumerator can easily tell, through further conversation, whether the respondent actually has no opinion, or not enough knowledge for him to form an opinion. There certainly are people in the agricultural areas of our country who do not know enough about labor unions or labor-capital strife to have a well-based opinion of it; or comprehend fully the urban housewife's food problem during periods of shortages or rationing.

Make the form easy to read and fill out and summarize. Have places for checks wherever possible, rather than words even as short as YES and NO. All spaces for such checks placed in vertical columns along an edge of the sheet make summarization easier. When you expect information written in, provide adequate space for it. One line for writing in a reason for doing something will keep the individual from explaining what he really means. When an interviewer is taking the record, he needs less space as he will put down just the key ideas.

Allow space for the name and address if you want them, and in most cases you will.

The problem of the questionnaire, the wording of the questions, and the asking of the questions, is just beginning to get its share of research. I discussed the sampling error in the discussion on sampling; there has been a great deal of research on sampling and we know pretty well what statistical error we can expect from sample surveys, and how to hold this error to a minimum.

The non-sampling error, caused by inadequate questionnaires and the handling of these questionnaires, is one that, as yet, cannot be measured. We do as well as we can to hold it to a minimum, but since it cannot be measured, we are never sure that it is at a minimum.

During the war, the Navy and Army did research on research methodology, and George Hausknecht, now a research consultant, gave me the following ideas that they had found out during their work:

In the listing of items, about which respondents should give a value or indicate an interest, or a knowledge of, error is likely to creep in, regardless of how you list them. When items are grouped according to subject matter, respondents are likely not to analyze the items separately, but to rate some according to the way they rated neighboring items. If items are listed in a random order, then the respondents are likely to relate items which are next to each other, even though they actually bear no relationship to each other, and rate them similarly. As far as I know, research has not proved which way is the best way.

Mr. Hausknecht said, also, that listing items and putting the boxes for response checks in a table arrangement causes error, as the respondents are likely to pattern their responses in ways not justified by the true responses. Rather, each item should be asked as a separate question, and the boxes or blanks left for the answers should follow each item and be separate from the other places for responses.

Another point he made emphasized something we have recognized for a long time, but probably have not paid it enough attention. Instructions for answers should be repeated again and again, and preferably with each question. If the responses desired are in answer to multiple-choice questions, then following each question should be the instruction to "check only one response," or to "check as many as apply," whichever the instruction is.

The following is an example of what can happen when a question is asked in two different ways:^{1/} When the question was stated "Tuberculosis may be inherited" with a choice of a True or False answer, between 45 and 47 percent of several thousand respondents answered True. When the question was started, "Which one of the following conditions is hereditary? (a) Tuberculosis, (b) color blindness, (c) syphilis," between 10 and 14 percent answered that tuberculosis was inheritable.

Another illustration of how wording may appear to distort results is afforded by two questions having to do with the age at which whooping cough is most dangerous. The questions were worded as follows:

44. Whooping cough is most dangerous to:
(a) Youth; (b) Infants below 2 years;
(c) Adults (in test 2).

48. Whooping cough is most dangerous to:
(a) Babies; (b) Children; (c) Youths;
(d) Adults (test 4).

^{1/} Source: Derryberry, Mayhew; Weisman, Arthur; and Caswell, George. What the Public Knows About Health. Natl. Institute of Health, U. S. Public Health Serv. American Museum of Health, Inc., 1790 Broadway, New York, December 1942.

The percentages of correct response in New York are 68 and 51, respectively, and in San Francisco, 70 and 54. However, if, in the answers to item 48, the percentage that responded "children" be added to that for the answer "babies" -- there is then no conflict in the results of the two items in either the New York or the San Francisco material.

Pretesting the Questionnaire:

No question should ever be used in the actual collection of data or information of any kind without its first being tried out to see if it is reliable and valid.

It is invariably a good idea to have several other people in your own office read over your questions to see if they are clear. However, this is not enough to make you sure that they will be clear to the people who will answer them. This is even more important for mail questionnaires than for personal-interview ones.

To pretest, go to several people typical of those who will later furnish you with the actual data, and have them answer the questions. If it means driving out to a county or across your county to find these people, and a day or so of work to do so, it is more than worth the time. There is no maximum number of people with whom you should pretest the questions. By all means, try them out with two or three to begin with. Then if certain questions are difficult to understand or answer, change them and try them out with some more. Continue this until there is no more question about their clearness.

Discussion IX. TABULATION AND SUMMARIZATION

Once data have been collected on a number of questionnaires, you are faced with the job of summarizing these data to give a picture of the group. As much time should be allowed for this process as was allowed for planning and making up the record forms and for collecting the information.

Calculations Needed:

Tabulation itself is purely a technique to use in getting accurate and complete statistical answers to the questions you had in mind in making up the questionnaire. You will remember that I said, when you make up a questionnaire, you first say what you want to know, and then ask a question that will get you that information.

Before the process of tabulation can be planned, you need to decide what kind of answers you want the data to give you. You look at each question, and decide what you want from the answers. (1) For example, you want to know how many said YES to a given question, how many are in different age classifications, and so forth. That is one type of answer available from questionnaire tabulations. (2) Another type is illustrated by your wanting to know what percentage of the people, or what proportion of the people, answered YES to a question, or had electricity, or fed certain kinds of feed to their calves. (3) Another type of answer is that you want to know the average number of acres in the farms studied, the average age of the homemakers, the average number of years in 4-H work, the average number of rooms per house. (4) A fourth type of answer is that you want to know what proportion of certain classes of your respondents give certain answers to given questions. For example, the proportion of the farmers with more than 160 acres of farm land who adopt certain practices, and the proportion of farmers with 160 acres or less who do; the proportion of women under 35 who come to meetings, and the proportion of women 35 and over who come to meetings.

Sometimes answers to questions are combined to give you the final answer you want. In getting at Extension contacts, we sometimes are not so interested in the number who have had certain specific kinds of contacts, as we are in the number who had any kind of contact, but in the questionnaire we usually get the information by asking about specific contacts. When we do this, we cannot just add together the number who attended an Extension meeting, the number who received an Extension bulletin, the number who listen to the county agent over the radio, and so forth, because many of these people will be duplicated in the different answers. Instead, we have to take off the questionnaire something to show which respondents had one or more Extension contacts, and add these tallies or checks.

As soon as we decide that we want our answers to be in terms of percentages, proportions, or averages, we need to get two totals -- the number who gave the answer in which we are interested, and the number we want to use as a base for the percentage or average -- in other words, the number by which we will divide.

It is to your advantage to know what kind of calculations can be planned for. There are three kinds of plain totals, or sums, of the answers given: One would be the number of checks in some certain blank; a second would be the sum of the numerical entries in given blanks; and a third would be the number who made certain kinds of answers -- the number who had three rooms, the number who had four rooms, the number who had five rooms, and so forth.

Measures of central tendency are those calculations which tell you the mean, the median, and the mode. The arithmetic mean (the ordinary average) is most commonly used, and is usually satisfactory, although in many cases where there are a few extremes, the median is better. The median is the middle value -- as many records are above it as below it. A family of 20 in one group can make the average for the group go high, whereas the median size of family for this group may be the same as for the other groups. Unusual cases of extremes, which by chance may fall in one group and not another, affect means, but not medians.

In many cases, the measure of central tendency, the mean or the median, is inadequate for telling the complete story of the data. Not only is it of interest to know that the houses in an area average 4.5 rooms; it is of great interest to know that 10 percent have only 1 room, that 10 percent have 10 or more rooms, and that 80 percent have other numbers of rooms. In deciding to calculate an average, think over the purpose for which you want the data and consider whether the distribution, or a measure of dispersion, is also necessary. (Example: Median schooling of rural people.)

Percentages are the other common calculation made from summarized data. Rather than say that 450 out of 980 people live on hard-surfaced roads, it is usually reported that 46 percent of the people live on hard-surfaced roads. Percentages are used in preference to raw figures for two reasons: To reduce the figures down to a size more easily understood by most people. Four hundred and fifty out of nine hundred and eighty doesn't mean very much to most of us. The other reason is to make possible the comparison of two groups of different size. If 450 out of 980 people in one county live on hard-surfaced roads and 680 out of 1,132 people in another county do, in which county does the larger proportion of people live on hard-surfaced roads? Similarly, it makes comparing of small groups clearer -- 25 out of 65 people compared to 13 out of 35 is not as clear as comparing 38 percent of 65 people with 37 percent of 35 people.

In calculating percentages, the base you use for calculations makes a great deal of difference in the interpretation. Ask yourself, for each case, whether you are interested in the percentage of your sample, the percentage of those who answered the question, or the percentage of those who answered a certain way to another question. When there are several "Not reported's" the percentage answering a certain way to a question will be different when based on all records or only on those reporting. Let us say that out of 100 records, 10 did not report to a certain question, 70 answered YES and 20 answered NO. Seventy percent of all records answered YES but 77 percent of those "answering the question" answered YES. Which is the better answer? Since we don't know whether the "Not reported's" would have answered YES or NO, it is essential that you either include in your analysis the fact that 10 percent didn't answer, or omit these records from the base.

It is not a good idea to use percentages if increasing the size of the figure makes the figures look more significant than they really are. If working with just one group, you find that 27 out of 35 people call on the county agent, this means that 77 percent of the people call on the county agent. If 77 out of 100 looks more significant than 27 out of 35, just through sheer size of figures, do not use it. Converting totals into percentages does not make the results more significant.

Only the more simple statistical calculations have been considered here. It is seldom that in our work we need the more complicated ones such as coefficients of correlation, and various statistical tests.

Tabulation

Tabulation consists of listing or grouping the answers from all members of your sample, or from all members of groups within the sample, so that summaries of the answers can be obtained.

Score or Tally Sheet

This is one of the more simple methods of tabulating. It consists of making a card or sheet on which are listed the possible answers down the side. Then across from each possible answer, one tally mark is recorded for each record which gave that answer. On this tally card may also be indicated the special sub-groups to which each record belongs. For instance, if you want to separate the farm from the nonfarm records and tabulate them separately so that those two classes can be compared, a separate column on the tally card can be designated for the tally marks from the farm records and another column designated for the tally marks from the nonfarm records. The results for the two groups can be added together, after the tallying is completed, to get a total for the whole sample.

This method of tabulating by a tally sheet is most valuable for tabulating a small number of short records. The greatest disadvantage there is the likelihood of error and difficulty in checking for error.

There is no identification of any tally mark to indicate from which record it comes; therefore, if the tabulator is interrupted in the midst of his job, he may not know where he was stopped, and will need to re-tally all his records to find out whether he had put down a certain tally or not. If, when checking the tallies, the number of tally marks doesn't agree with the number of records, he cannot tell where the error was made, and continuous checking is necessary until several counts agree.

A tally sheet is used in connection with the tabulation of big studies for two purposes: (1) To take off preliminary, unchecked data for immediate reporting, and to get a hint of the story that will be shown. These data are not published. (2) To get an estimate of the number of records showing certain results to determine classifications for comparison. If lower-income families are to be compared with higher-income families, a quick tally will show you where the break comes between the low-income levels and the higher ones.

Hand-sorting

This method consists of counting without tallying. A card or sheet is made up similar to the tally sheet, except that space need be left only for the total of each group and answer. If the records are to be grouped (farm vs. nonfarm), they are sorted into piles of each group before counting starts. This method is even more liable to error than the tally method; it is best used when the record form is a small card which is easily handled. Hand-sorting is often used for the same purposes as the tally-sheet method.

Tabulation Sheet

This method combines the tally-sheet and hand-sorting methods, and allows for the reproducing of the whole questionnaire onto the tabulation sheet. It is not as fast as either of the two described methods but is a great deal more accurate and much better for permanent record. Its accuracy can be as perfect as human work ever is; by checking and rechecking of the tabulation, the tabulation can become even more accurate than the collection of the data was.

For this tabulation-sheet method, large sheets are set up with the questions and the possible answers either across the top of the sheet, or down the left-hand side, whichever seems more convenient. Across the top is usually more convenient when tabulating is done on the typewriter because typing is easier and more natural going across the page. In this case, each record is allowed one line across the page. At the extreme left-hand edge of the page is entered either the record number or the person's name (some definitely identifying symbol), and then the answers are entered as x's or figures or words in the proper columns (question answers).

The way in which you allow columns for the answers is determined by your choice of using more paper, or simplifying the adding or summarizing. For instance, if the possible answers are YES, NO, DON'T KNOW, and NOT REPORTED, you may allow a column for each of these four possible answers and put the check in the proper column; or you may leave one column for the question, and enter either YES, NO, DON'T KNOW (DK), or NOT REPORTED (NR) in the column. Whichever way is used, when all the records are tabulated you will have a complete picture of all the records taken, in a form in which they can be easily summarized.

To separate the data into groups for comparison, the records must be classified before tabulation starts. This indicates the necessity of knowing, before you start tabulating, what comparisons you wish to make -- what kinds of results you will want. The records can be broken down into any number of sub-groups, and the various sub-group totals added together in various combinations to get various group totals. This is, admittedly, a cumbersome method of getting group comparisons and causal relations, but it is effective and possible when electric machine tabulation is impossible.

Electric Machine Tabulation

There are various companies making various types of machines for tabulating and summarizing answers from record forms or questionnaires. Some questionnaires, set up for certain machines, are marked with a certain kind of pencil so the forms themselves can be run through the machine. The pencil marks make electric contact for a counting of the answers. The most common kinds of machines are IBM and Remington-Rand. For our purposes, either can be used. For these machines, every answer on the record form is reduced to a numerical or alphabetical code, then these are punched onto a small card. This results in a small card carrying the whole story of one questionnaire on it. These cards are run through sorting machines which sort out the answers to any questions we want, giving us the number answering each type of answer. Or the tabulating machine totals the entries, giving us the sum on all the questionnaires, and the number reporting the answers.

If the questionnaires are set up so that the answers checked are automatically assigned numerical values, the work and time involved in electric-machine tabulation is less than for typewriter tabulation, and there is less chance for human error. The great advantage, however, of electric-machine tabulation is the increased freedom in manipulation of the cards. Groups do not have to be all predetermined. If an interesting causal relationship is indicated in early summaries of the data, the cards can be sorted on the machines immediately, with no necessity for a complete retabulation as is necessary with typewriter tabulation.

Free-Answer Information

This was referred to in the discussion on coding. Many times it is impossible to classify free answers until they are all available and all can be seen and studied as a unit. In these cases, the free answers are typed out on large sheets -- the more answers on a single sheet the better. Then someone, preferably the professional person in charge of the study will need to study them, breaking them down into a few large classifications or types of answers, and then into groups and sub-groups within these types, until we can begin to say that a certain number of the respondents gave a certain answer, a certain number another answer, and so forth. If different groups of people in your sample are to be compared as to how they answered these free-answer questions, the free answers will then need to be coded and retabulated according to the group of the sample to which they belong. When electric-machine tabulation is used, this classifying of free-answer questions has to be done before the card-punching is done.

Discussion X. INTERPRETATION OF DATA

Forms in Which Data Appear

Data usually appear in tables or on charts as the number of people or things that have certain characteristics or who do something; as the percentage who have these characteristics or who do something (remember that percentage means the number per 100 -- that is, out of every 100 persons or things, that number has the characteristic or does the thing indicated); or as the average number who have certain characteristics or who do something (an average is the number per the number reporting).

The data are presented in these ways in order to give you the picture of the group described; the characteristics of any one individual is of no importance in a statistical or sampling survey. Combining answers of all individuals gives us the picture in general of the group. If the sample is chosen to represent everyone in the total group, then the data from the combined individuals surveyed give you the picture of the total group or population.

Reading a Table or Chart

There are a few suggestions for reading of a chart or table which may help the person to whom an array of statistics is only bewildering. Read the headings across the top to see into what groups the data are classified or are described; then read down the left side to see what these items are. Then decide if you want to see how the data differ from group to group as classified across the top, or how they differ from group to group as classified down the left-hand side. This determines whether you will read the table across the lines or down the columns.

Significance of the Findings

Whether the size of any given figure is significant -- whether it is considered a large percentage, or a small average -- this decision can be made only by the person who has knowledge of the field in which the data lie. A statistical clerk may hand you a figure which he has calculated -- an accurate figure based on reliable data. The significance of that figure can be determined only by you, if you know anything about the field in which it lies. Seventy-five percent of the women in a given community belong to a home demonstration club; that figure would be a significantly large percentage because past experience and research have shown us that it is unusual for that many women in any community to belong to that kind of an organization. The interpretation can be carried even farther -- was the figure large for the given community? Might it not have been a small, closely-knit, homogeneous community in which practically all the women always belong to the same organizations; in which case 75 percent may not be large; it may even be undesirably small.

Only the person who knows the subject-matter field involved can determine whether he can say, "Only 50 percent of the men adopted a certain practice," or "Already, 50 percent of the men have adopted a certain practice." Only he can know whether 50 percent is failure or an attainment of which to be proud.

Statistical Significance

Mention has already been made of statistical error in sample data -- the error which results from the data being collected from only some of the people and not from all of them. Appreciation of this statistical error enters into an interpretation of the significance of results. When comparing two figures from the same set of data obtained in a sample survey, we know that they will differ some because of this statistical error. But what we must know is how wide is that error? How much difference must there exist between two figures before we can say that one is significantly larger than the other?

If 35 percent of the people knew some piece of information before the lesson, and 46 percent knew it afterwards, is that a significant difference? Might not a different percentage result by chance from the after-test and from the before-test? Formulas similar to the standard-error formula mentioned in the discussion on Sampling can be used to test whether the difference between such percentages (or between averages) is significant. In using the formulas, be sure that the sample from which you obtained your data is a representative sample of the "population."

In few cases can we feel sure that a small difference between percentages is significant. Using a statistical formula, we find that with 100 records, the difference between two percentages of 41 percent and 35 percent is not significant. With a small sample I hesitate to say that a difference in percentage of less than 10 percent between two percentages indicates such a difference actually exists in the "population" from which we selected the sample. If, however, I find that all evidence I have collected points in the same direction, if all percentages and averages show the change of behavior in a given direction, each individual difference in percentages or averages may be significant even if small.

In considering that our data, when collected from a sample, give us only an approximation to the true result, we can see that the retaining of decimal places in percentages and averages which contain whole numbers may give a false impression of precision not always justified by the size of the sample.

Consideration of the Important Factors

A problem in interpretation of data is deciding what relationship between sets of factors needs emphasizing or to be brought to light. In some cases, there is a relationship between two sets of factors which is not a cause-effect relationship; in other cases, there is a cause-effect relationship. In the first, we need to ask ourselves on which classifi-

cation we wish to base our findings. For example, if you have your records classified as to education and as to kind of farm, we may wish either to show what percentage of the people with each level of education live on the different kinds of farms; or to show what percentage of the people living on the different kinds of farms have the different levels of education.

A typical cause-effect relationship may be illustrated by the relationship between the reading of bulletins and the level of education. The latter is obviously the causal factor, the first the result or effect factor. Ordinarily, we would sort the records according to level of education, and calculate what percentage of each of these classes reads or doesn't read the bulletins. However, in your analysis you may wish to compare the readers and nonreaders as to level of education, in which case the cause-effect relationship goes out of the picture, and the sorting is done by readers and nonreaders, and interpretation is made from that viewpoint.

Interpretation of the data depends on the way the data have been classified and related to each other, and tables should show clearly how that has been done.

Interpretation of the Data in General:

Interpretation of data is seldom limited to interpreting data you yourself have collected. Most often, you are interpreting data someone else has collected, tabulated, summarized, and analyzed. Dr. G. H. Aull of South Carolina, in a talk to a subject-matter specialists' workshop in South Carolina, gave some points to keep in mind when interpreting others' economic and social data. These points certainly hold true when interpreting your own, too. Following are these points:

- Where did the data come from?
- What was the purpose in collecting the data?
- How were the data obtained?
- Do we have all the facts?
- Look at the definitions.

Dr. Aull also called attention to limitations to data which should be watched for. The points he brought out in regard to this subject were:

- Limits of definitions.
- Sampling method or size of sample.
- Conditions prevailing when data were taken.
- Interpretation put on by investigator (Miss Hall's interview).

Discussion XI. WRITING YOUR ANNUAL REPORT

Once a year, every Extension worker has the opportunity to summarize his daily and periodic evaluations of his work. There are three different attitudes toward this yearly activity -- that it is a chore that must be done, that it is the one time the worker evaluates his activities, and that it is, as I expressed it above, the opportunity to summarize one's evaluation of his work. With the first attitude, it is psychologically painful; with the second attitude, it is extremely difficult; with the third attitude, it may still be a lot of work but it should be neither painful nor extremely difficult. You will remember Miss Collings' paper that I assigned you for reading earlier; it shows how evaluation is done during the year as teaching plans are made and teaching is carried on.

Before we can consider accurately how to prepare the annual report, we need to know the purposes of this report -- in other words, its objectives. There is hardly anything that we can plan, do, or consider, without going first through that early step in the evaluation procedure -- what are the objectives?

The purposes of the report have been summarized as follows:

1. To inventory past year's efforts and accomplishments in order to plan more effectively for another year.
2. To improve oneself professionally through self-appraisal.
3. To provide supervisors with basic information required for appraising the work accomplished.
4. To record the year's work in shape for ready reference by oneself or others (especially successors).
5. To account to the tax-paying public for what Extension has accomplished.
6. To fulfill one's obligation of the job as defined by the Smith-Lever law.

These six varied objectives immediately call to attention that the "population" must be considered. In other words, different kinds of people are going to read and use the report, and it has to be written for them all. Six kinds of readers have been listed for these reports.

1. The worker himself (most important).
2. County planning group.

3. Fellow workers and successor.
4. Supervisors.
5. The Government - State and Federal.
6. The public.

However, these six kinds of readers all are looking for one thing in common: What happened as a result of what was done. And what happened and what was done in line with objectives or purposes. The different readers may have more interest in one part than another -- the public is probably most interested in results; county planning groups in the relationship between objectives and results; fellow workers, successor, and supervisors in all three: Objectives - methods - results. Certainly, the worker himself is vitally interested in the relationship between the three.

Provision has been made for the county workers to report their work in two ways: The statistical report and the narrative report. They each have their usefulness.

Statistical Report:

Most of this report is either only measurement, if you want to look at it that way, or is limited evaluation at the opportunity level. The number of meetings held is purely measurement -- an attendance of 1,268 people at a series of meetings does give you the number of people attracted to the meetings -- the number who had the opportunity to receive information. Part of the questions in the annual statistical report cover the number of people or families or groups helped with certain problems. These actually are evaluations of certain programs if the numbers entered in the report are determined reliables and with validity. The same figures, however, give no clue to quality of the help.

These statements are in no way a criticism of the statistical report; they are part of an objective review of what is contained in it. To the degree that the statistics give the picture of the work of the agent, to that degree they can be used for that purpose. Mostly, the statistical report is a measurement of what the agent has done during the year, not how well he has done it. Within the limits of the items contained in the report, it tells of methods -- not objectives and, to a large degree, not results.

Narrative Report:

The narrative report provides the opportunity for the agent to fill in everything that the statistical report limits him from telling. He tells what, how, and the results. He tells with whom, and what they are like. He can explain why and why not.

Uses Made of Annual Reports in Federal Office:

Where received.
How handled.
What is done with them.
Summaries and analyses:
For what used.

Application of Evaluation Outline to Narrative Report:

The attached basic outline for the narrative report can be applied to the write-up of a single activity, meeting, or project, to a complete phase of the program, to the complete program of a given worker, or to the total county program. It provides the opportunity to tie in each activity or phase of the program to the total, or the work of any given worker into the total work done on the total program.

It is quite obvious that this outline provides for the evaluation of the work of an Extension worker and not for a story of his activities. It is aimed at describing the results, with the methods and activities only as tools.

I was interested to run into the following sentence in a county agricultural agent's annual report: "Since no goals were set for work to be accomplished during the year, no measure has been made of the accomplishments ...". It is obvious that he considers his results in terms of his objectives.

Discussion XII. USE OF EVALUATION FINDINGS AND PRESENTATION OF DATA

The main use to be made of what you learn through evaluation is applying it to your future work. A report of the findings or information is purely a tool, and not an objective. A report is written for one or both of two reasons: To clarify the thinking of the evaluator, and second, to present it to other persons. When I give a test in this class, I find out what I failed to teach in the few days prior to the test, and what I did teach; I don't write a report of what the test showed me, but I apply my findings in two ways: I try to make up for my failure in the days ahead, and the next time I teach the course, I try to present those ideas that I failed to present well this year in a different way to see if they will go over better. I do, however, make a sort of report of the test results when I report grades to the graduate school at the end of the session.

County agents do the same thing. Here's a quotation from a county extension agent's 1950 report: "The dormant or not-so-active town (4-H) committees have not been able to get the local 4-H program to click. This has been due in part, to (1) poor choice of chairman; (2) lack of initiative on members' parts, (3) not enough follow-up by county club agents. The county club agents plan to spend more time with these dormant committees during the coming year."

In contrast, here is an optimistic result of his evaluation, "Despite the low enrollment, this is the first year that I feel encouraged about the outlook for 4-H garden club work. And I am encouraged because (1) Town committees are finding leaders and members, and (2) Local leaders are very much interested in their club program."

The important thing is that the evaluation process has not served its purpose unless the conclusion drawn from it has entered the continuing plans of the worker. If it is found out that present activities of the worker fail in reaching certain objectives or certain defined "populations," the worker has not improved himself professionally or the value of his work if he does not attempt to modify his activities so that they might or will reach the objectives or the "population."

We are interested in how to present findings or data when they are to be presented to other persons. If they are presented in a talk to a group of local leaders or an advisory board, they might not be published, but the rules for presenting findings apply in most cases, to the talk as well as to the publication.

The way in which findings are presented is determined by the audience or reader. If they lack the training or background for interpreting statistical tables or discussions but are supposed to use the findings in their work, the findings should be entirely interpreted for them, and presented to them in terms of activities or recommendations that they can

understand. If the purpose of the presentation is to provide statistical proof for a decision or recommendation, the statistics need to be included in the presentation or publication. If the audience or readers are interested in the procedure involved in collecting the data, that should be included; if the inclusion of such details would detract from the importance of the findings or confuse certain types of audience or readers, it would be limited to the barest essentials. In many cases, the findings from surveys have to be presented in more than one way to meet the needs of several kinds of audiences or readers.

The Report:

In general, the following points should be contained in any report or series of reports of a survey:

1. The purpose of the survey.
2. For whom and by whom the survey was conducted.
3. General description of the universe covered.
4. The size and nature of the sample and description of any weighting methods.
5. The time the field work was done.
6. Whether personal or mail interviews were used.
7. Adequate description of field staff and any control methods used.
8. Questionnaire and findings.
9. Bases of percentages.
10. Distribution of interviews.
11. Author of the report.
12. Date of publication of report.
13. Where published.

Presentation of Data:

Data may be presented in five principle ways:

1. They may be presented entirely in word descriptions, with no use of figures.

2. Incorporated in the text. -- The data you gathered, the calculations you made, may be used as part of the narrative form of a report. The advantages of presenting data in this way are that the interpretation of the data accompanies them, and that many people do not read tables and charts easily or understandingly.

3. Tabular form. -- This means putting the data into tables or lists. Tables may be of two kinds: Large, comprehensive tables, used mainly for reference from the text, and which may be put in the Appendix at the end of the report; and smaller, simple tables which are part of the text and the interpretation of the figures close to it. I mentioned in an earlier talk the essential characteristics of tables.

4. Semi-tabular arrangement. -- This is a combination of the text and tabular plans. When only a few figures are to be used in a discussion, the text may be broken and the data listed.

5. Graphic presentation. -- There are innumerable kinds of charts, graphs, maps, and diagrams which can be used to present data. Common kinds used in reporting the type of data we are most likely to gather are:

a. The simple curve diagrams. Curves are used primarily for picturing time series, and for showing frequency distributions. The main point to be remembered in using the line chart is that it is not used unless the spaces on the horizontal axis (base line) could be broken up into smaller and smaller divisions; unless at any point on the line, chosen at random, you could be justified in reading the chart.

b. Bar charts. These are used with the bars either horizontal or vertical, and are used mostly for comparing magnitudes. The difference in length of bars on a chart is, for most people, easier to see than the difference between figures. Bars are also used advantageously for showing component parts; that is the parts of a total.

c. Pie diagrams. I am mentioning this type because it is often used, and usually quite effectively. It shows either of the things that the bar chart does -- comparison of magnitude, or component parts. It should not be used in preference to a bar chart, except to give variety to the presentation. Bar charts, involving only one dimension, are easier to read.

d. Pictorial diagrams. These are more interesting to read than the ordinary line or bar chart. Usually we don't have facilities to make them, but when we do, they offer variety. The main rule to remember is to increase the number of pictures to indicate increase in magnitude, not increase in size. Difference in size of pictures is very hard to interpret. Other rules governing pictorial charts are: Symbols should be self-explanatory; pictorial charts compare approximate quantities, not minute details; and only comparisons should be charted, not isolated statements.

e. Statistical maps. These are graphic devices which show quantitative information on a geographical basis. Hatched or shaded maps undertake to show for each geographical area the magnitude of the phenomenon which is being studied. When the object is to show the geographical distribution of occurrences, the dot map should be used.

Writing the Text:

In planning your report, and the text of it, try to think through the story that you want to tell your readers, then write that story in your own words.

Some of Mrs. Amy Cowing's teachings on readable writing seem so fitting to our problem of writing up study findings that I have copied some of them here:

When you plan your report and write it:

"SELECT

"What you want to teach. Select the ideas you want to get over to your readers. Decide: Is information suitable? Does it meet a need? Is it timely? Is it of current interest? Does it apply to your area? Is information practical? Can people use information? Can they afford to use it?

"Who you want to teach. Who are the people you want to reach? What are their problems, interests, and educational levels. Do your readers have the environment and capacity to make use of information?

"Why you are putting this out. What is the purpose of the publication? What do you want it to accomplish? Do you want to stimulate interest in a program, or do you want to influence people to act -- to adopt a new practice?

"SIFT facts

"Sift essential facts necessary to give information clearly. Screen out difficult concepts that are beyond readers' experience or understanding. Give laymen an appreciation of subject rather than a detailed explanation. Express highlights; don't impress lay reader with all you know. Don't document everything.

"SORT facts

"Arrange facts in logical order. Give information so reader can follow it step by step. Help reader dig for facts; set out important points in 1-2-3 order. Guide reader with snappy subheads.

"SELL facts

"Get reader to read information and to act on it.

"SAY facts in

"Short sentences -- the shorter the better. Give only one idea in each sentence; express two related ideas clearly in two independent clauses separated by a semicolon (like this compound sentence). Avoid 'tapeworm' sentences; complicated, complex sentences that run on and on with many 'which' clauses.

"Short Words -- the shorter the better. Use the simplest word that carries your meaning. Use concrete picture words. Translate abstract words into concrete. Explain difficult, technical words that have no simpler synonyms.

"Personal references -- the more the better. Sprinkle generously with human-interest words, such as 'we, you, mother, brother, folks' and other 'people' words. Person-to-person style of writing 'talks' to reader in active voice -- the way we speak to each other. Cookbook style implies 'you'; this is one of the best ways to give information simply."

Summary:

To summarize the presentation of study findings, remember that data should never be collected for data's sake; neither should data be presented for data's sake. Data are collected to answer a question in your mind; when they are interpreted, they give you an answer to that question. The report should be a summary of those answers.

